Controlling Elevation and Azimuth Beamwidths with Concentric Circular Microphone Arrays

Rajib Sharma*, Israel Cohen, and Baruch Berdugo

***Abstract*—Solutions for frequency-invariant or constant-beamwidth beamforming with concentric circular arrays (CCAs) generally control only the azimuth-beampattern. This work introduces a beamforming methodology, which instead of designing a fixed beamwidth across the spectrum, maximizes the directivity-factor under the constraint that the beamwidth is greater than or equal to a pre-defined target-beamwidth, in both the elevation and the azimuth. The principal idea lies in dynamically weighting the microphones of the CCA based on two factors - (i) their corresponding ring number or radius, and (ii) their alignment with the signal-path. The parameters controlling the weights are determined using a modified gradient-descent algorithm. Experimental results show the flexibility and advantage of the proposed methodology over state-of-the-art methods.**

***Index Terms*—Concentric circular array, constant-beamwidth beamforming, microphone array, frequency-invariant beamforming, directivity-factor, white-noise-gain.**

I. Introduction

THE objective of beamforming is to extract a signal-of-interest from the detected/sensed data, based on its direction-of-arrival (DOA) with respect to the sensor-array [1]–[13]. The DOA, given by the elevation angle, θ_{d}, and the azimuth angle, ϕ_{d}, introduces relative delays in the sensors, which are represented in the frequency-domain or phase-form by the steering-vector. A filter, called the beamformer, is utilized to retrieve the signal-of-interest corresponding to the steering-vector, amidst other interferences and sources of noise. In general, as the frequency increases, the 3-D beampattern, directed towards the DOA, becomes narrower (in both the elevation and the azimuth). This phenomenon is undesirable: if the beampattern is too narrow, then, any minor misestimation of the DOA may result in distortion or loss of information in the beamformer output. Of course, having too-wide a beamwidth (as in low frequencies) is also undesirable, as it could introduce interference(s) from other directions. Consider now that the signal-of-interest is produced by some complex natural process, as in the case of the speech signal. The speech signal has a multitude of information embedded across its entire spectrum, such as the message-content; the identity, emotion, health, etc. of the speaker; and even the acoustic conditions [14], [15]. As such, in the case of the speech signal, the variation in the beamformer characteristics with respect to frequency could be a stark impediment for the target application depending on the beamformer output. For such reasons, the domain of frequency-invariant or constant-beamwidth beamforming is an active area of research [16]–[24].

Frequency-invariant beamforming is principally focussed on replicating a desired beampattern in the azimuth plane for a limited pre-defined frequency range. A more refined goal is called constant-beamwidth beamforming, wherein a constant pre-defined azimuth-beamwidth is sought for the desired frequency range. In the case of a uniform linear array (ULA), which is a 1-D array, the steering-vector has a Vandermonde structure, which simplifies mathematical analysis, and hence has resulted in a multitude of analytical algorithms in such endeavors [25], [26]. However, a ULA cannot scan the entire azimuth plane - for two signals with the same θ_{d}, the ULA cannot distinguish between them if they are arriving from two azimuth directions, $\pm\,\phi_{\mathrm{d}}$, symmetric with respect to the end-fire direction. A circular array, which is a 2-D array, is devoid of such a limitation, and hence is utilized in many practical applications [19], [27]–[30]. By placing the sensors in multiple concentric circular rings, a more versatile design is achieved, known as the concentric circular array (CCA) [16], [17], [31], [32].

Obtaining an analytical solution to control the elevation-beamwidth in the case of a planar array is challenging. This may be the reason why known solutions of frequency-invariant or constant-beamwidth beamforming do not consider the elevation-beampattern in the case of CCAs. Consider now the fact that lower frequencies naturally have wider beamwidths. As we will later observe in this work, forcing a substantially lower desired beamwidth at low frequencies may result in an arguably impractical solution. Again, as we will explore in this work, the elevation and azimuth beamwidths share a complex inverse relationship. As such, forcing a strict beamwidth requirement on the azimuth plane might result in undesirable beampattern characteristics in the elevation plane. Therefore, in this work, we relax the beamwidth criteria and strive to obtain a beamwidth wider than a certain pre-defined desired beamwidth in both the planes (instead of a fixed beamwidth in either plane).

In most existing works on frequency-invariant or

This research was supported by the Israel Science Foundation (grant no. 576/16) and the ISF-NSFC joint research program (grant No. 2514/17). I. Cohen and B. Berdugo are with the Andrew and Erna Viterby Faculty of Electrical Engineering, Technion – Israel Institute of Technology, Technion City, Haifa – 3200003, Israel. Email(s): icohen@ee.technion.ac.il, bbaruch@technion.ac.il.

R. Sharma (*Corresponding author) was a Postdoctoral Fellow at Technion. Email: rajibd2k@yahoo.com.

constant-beamwidth beamforming, the signal-of-interest is confined to the horizontal plane ($\theta_{\rm d} = 90^o$), which simplifies the analysis. It is expected that the solution derived would be applicable for any arbitrary $\theta_{\rm d}$. However, as we will explore in this work, solutions derived under this condition may not generalize well to more practical scenarios where the signal-of-interest impinges on the CCA at an arbitrary oblique elevation angle ($\theta_{\rm d} \nsim 90^o$). Again, in most works on frequency-invariant or constant-beamwidth beamforming, a central sensor is generally not included in the CCA [16]–[21], [33]. Similarly, constraints regarding the number of sensors in the different rings of the CCA exist in many solutions [16], [17], [20], [33]. Further, the effect on the directivity-factor (DF) and the white-noise-gain (WNG) is overlooked in many works, focussing solely on controlling the azimuth-beampattern.

At this juncture, we might clarify that this work is not devoted to studying the use-case of the central sensor, or the pros-and-cons of different ring designs or configurations, or related issues like the sidelobe levels [34]–[36]. This work is strictly devoted to deriving a practical frequency-invariant beamformer for any arbitrary CCA.

In our recent work, we derived an analytical solution to control the elevation-beamwidth of a CCA in the broadside scenario [33]. In this special scenario, the azimuth-beamwidth cannot be controlled and the number of sensors in each ring of the CCA must be regulated. Hence, the solution cannot be generalized. In this work, our goal is to obtain a solution for any arbitrary DOA and arbitrary CCA, which can control both the elevation and azimuth beamwidths, without neglecting the DF and WNG. For this purpose, we derive a numerical solution using a modified gradient-descent algorithm. We illustrate our solution for a CCA with a central microphone. To further ascertain that our proposed solution is applicable for any arbitrary CCA, we also implement it on a CCA which has fewer microphones than that required to avoid spatial aliasing.

The rest of this work is organized as follows: Sec. II presents the beamforming framework, CCA design, and performance metrics. Sec. III describes the proposed methodology. Sec. IV presents the results in comparison with other algorithms, and Sec. V summarizes this work.

II. Signal model, array design, and parameters

We consider a CCA consisting of P concentric rings lying on a horizontal plane, as shown in Fig. 1. The horizontal plane represents the X-Y plane in a 3-D Cartesian co-ordinate system, with (0,0,0) being the center of the CCA, as shown in Fig. 2. Let r_p denote the radius of the $p^{\rm th}$ ring, and M_p denote the number of microphones in it. Consider also a microphone at the center of the CCA, which may be considered as the $0^{\rm th}$ ring of radius, $r_0 = 0$, and $M_0 = 1$. Let $M = \sum_{p=0}^{P} M_p$ denote the total number of microphones in the array. Let ϕ_{p,m_p} denote the angular position (measured with respect to the X-axis) of the $m_p^{\rm th}$ sensor in the $p^{\rm th}$ ring. A far-field speech signal-of-interest, which is non-stationary and broadband in nature,

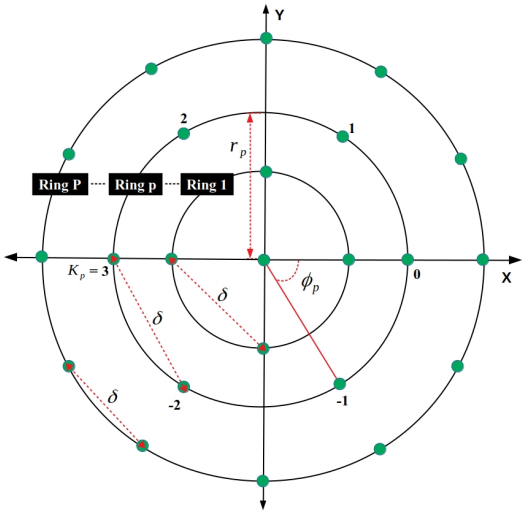

Fig. 1: **Diagrammatic representation (top-view) of a CCA. δ denotes the inter-sensor distance. r_p denotes the radius of the $p^{\rm th}$ ring in which the angular separation of the sensors is ϕ_p.**

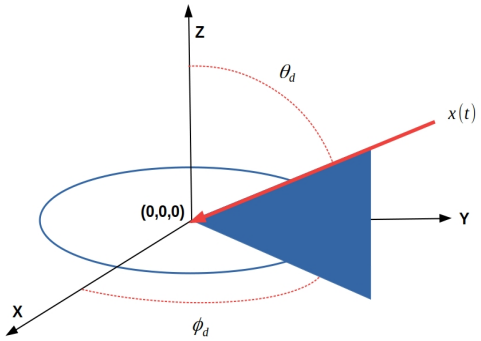

Fig. 2: **The circular co-ordinate system. A signal, $x(t)$, impinges at the center with a direction-of-arrival, $(\theta_{\rm d}, \phi_{\rm d})$. The center of the co-ordinate system may represent the center of a CCA.**

impinges on the CCA with the DOA, $(\theta_{\rm d}, \phi_{\rm d})$, as shown in Fig. 2.

Apart from the signal-of-interest, various other interfering signals may impinge on the CCA from other directions. Also, the microphones are contaminated with their thermal noise. The various signals are sensed/detected and sampled by the microphones at a sampling frequency of $F_s = 1/T_s = 16$ kHz. As such, during sampling, the signals are bandlimited to a frequency, $F_{\rm max} = F_s/2$. Thus, the minimum wavelength of the signal-of-interest is given by, $\lambda_{\rm min} = c/F_{\rm max}$, where $c = 340$ ms^{-1} is the velocity of propagation of sound in air at room temperature. Henceforth, the data sensed at any given microphone of any given ring of the CCA may be represented as

$$\begin{aligned} y_{p,m_p}(t) &= x_{p,m_p}(t) + v_{p,m_p}(t), \\ x_{p,m_p}(t) &= x(t - \tau_{p,m_p}), \\ \tau_{p,m_p} &= -F_s \frac{r_p}{c} \sin\theta_{\rm d} \cos(\phi_{\rm d} - \phi_{p,m_p}). \end{aligned} \tag{1}$$

In (1), t denotes the discrete-time index, and $x(t)$ the signal-of-interest impinging on the center of the array. Hence, $x_{p,m_p}(t)$ denotes the signal-of-interest sensed by

the corresponding microphone with a certain time-delay, τ_{p,m_p}, depending on the DOA. $v_{p,m_p}(t)$ denotes the total effect of interferences and noise on the microphone. The data received by the microphones is processed in frames that are represented in the frequency-domain. Thus, for a given microphone, we have

$$\begin{aligned} Y_{p,m_p}(f) &= X_{p,m_p}(f) + V_{p,m_p}(f), \\ X_{p,m_p}(f) &= e^{j2\pi f \frac{r_p}{c} \sin\theta_\mathrm{d} \cos(\phi_\mathrm{d} - \phi_{p,m_p})} X(f) \\ &= d_{p,m_p}(f, \theta_\mathrm{d}, \phi_\mathrm{d}) X(f), \\ d_{p,m_p}(f, \theta, \phi) &= e^{j2\pi f \frac{r_p}{c} \sin\theta \cos(\phi - \phi_{p,m_p})}. \end{aligned} \tag{2}$$

Hence, the data received across all the microphones of the CCA is given by

$$\begin{aligned} \mathbf{y}(f) &= [\mathbf{y}_0^T(f),\ \mathbf{y}_1^T(f),\ ...,\ \mathbf{y}_P^T(f)]^T, \\ &= \mathbf{x}(f) + \mathbf{v}(f), \\ \mathbf{x}(f) &= \mathbf{d}(f, \theta_\mathrm{d}, \phi_\mathrm{d}) X(f), \\ \mathbf{d}(f, \theta, \phi) &= [..., \mathbf{d}_p^T(f, \theta, \phi), ...]^T, \\ \mathbf{d}_p(f, \theta, \phi) &= [..., d_{p,m_p}(f, \theta, \phi), ...]^T. \end{aligned} \tag{3}$$

In the above equations, $\mathbf{d}(f, \theta, \phi)$ represents the steering-vector for the direction (θ, ϕ). The objective of beamforming is to apply an M-dimensional filter, $\mathbf{h}(f)$, on the data-vector, $\mathbf{y}(f)$, so as to obtain a signal, $Z(f) \approx X(f)$, which is an approximation of the signal-of-interest impinging on the array with the DOA, $(\theta_\mathrm{d}, \phi_\mathrm{d})$. In other words, the beamformer, $\mathbf{h}(f)$, must be steered in the look-direction, $(\theta_\mathrm{d}, \phi_\mathrm{d})$:

$$\begin{aligned} \mathbf{h}(f) &= [\mathbf{h}_0^T(f),\ \mathbf{h}_1^T(f),\ ...,\ \mathbf{h}_P^T(f)]^T, \\ \mathbf{h}_p(f) &= [..., h_{p,m_p}(f), ...]^T, \\ Z(f) &= \mathbf{h}^H(f)\mathbf{y}(f) = X_\mathrm{fd}(f) + V_\mathrm{rn}(f), \\ X_\mathrm{fd}(f) &= X(f)\mathbf{h}^H(f)\mathbf{d}(f, \theta_\mathrm{d}, \phi_\mathrm{d}), \\ V_\mathrm{rn}(f) &= \mathbf{h}^H(f)\mathbf{v}(f). \end{aligned} \tag{4}$$

In (4), $X_\mathrm{fd}(f)$ and $V_\mathrm{rn}(f)$ represent the filtered-desired and residual-noise signals, respectively. Ideally, one would want $V_\mathrm{rn}(f) = 0$, while obtaining a distortionless estimate, i.e., $X_\mathrm{fd}(f) = X(f) \implies \mathbf{h}^H(f)\mathbf{d}(f, \theta_\mathrm{d}, \phi_\mathrm{d}) = 1$.

A. Array design

To avoid spatial aliasing, the microphones in any ring of the CCA must be separated by an inter-sensor distance, δ, such that $\delta \leq \lambda_\mathrm{min}/2$. As such, a generic CCA design automatically implies that $\delta = \lambda_\mathrm{min}/2$. We denote this design as CCA-I. The number of microphones in any ring (except the central microphone), rounded off to the nearest even number, is given by

$$\begin{aligned} M_p &= \left\lfloor \pi / \sin^{-1}\left(\frac{\lambda_\mathrm{min}}{4r_p}\right) \right\rceil,\ p = 1, 2, ..., P, \\ M_p &\coloneqq \begin{cases} M_p, & \text{if } M_p \text{ is even} \\ M_p + 1, & \text{if } M_p \text{ is odd} \end{cases}. \end{aligned} \tag{5}$$

Thus, in CCA-I, the p^th ring of radius, r_p, consists of $M_p = 2K_p$ omnidirectional microphones. The microphones are uniformly distributed on the ring, with an angular separation of $\phi_p = 2\pi/M_p$. The angular positions of the microphones in the ring are given by

$$\begin{aligned} \phi_{p,m_p} &= m_p \phi_p = m_p \frac{2\pi}{M_p}, \\ m_p &= \begin{cases} 0, & p = 0 \\ -K_p + 1, ..., 0, ... K_p, & p = 1, 2, ..., P \end{cases}. \end{aligned} \tag{6}$$

While having the required number of sensors to avoid spatial aliasing is ideal, one may wish to obtain the necessary performance using fewer sensors. In that case, the inter-sensor distance will be $\delta > \lambda_\mathrm{min}/2$, and at higher frequencies spatial aliasing may occur. However, depending on the application, such a trade-off might be acceptable, particularly since at higher frequencies the DF is generally robust. Such a CCA design may be considered a sparse design. In this work, we will use the CCA-II design, where the number of microphones in each ring is $(1/4)^\mathrm{th}$ of that in the CCA-I design. The number of microphones in each ring of the CCA-II design is rounded off to the nearest even number and the microphones are uniformly separated, as per equations (5) and (6).

B. Metrics for optimization and evaluation

As is evident from (4), the objective of a beamformer, $\mathbf{h}(f)$, at any frequency, f, is to amplify the signal-of-interest in the look-direction, $(\theta_\mathrm{d}, \phi_\mathrm{d})$, while attenuating interfering signals from other directions. The beampattern of the beamformer represents this capability. It is defined as

$$\begin{aligned} \mathcal{B}(f, \theta, \phi) &= \mathbf{h}^H(f)\mathbf{d}(f, \theta, \phi), \\ &\theta \in [-\pi,\ \pi],\ \phi \in [-\pi,\ \pi]. \end{aligned} \tag{7}$$

Henceforth, $\mathcal{B}(f, \theta, \phi_\mathrm{d})$ and $\mathcal{B}(f, \theta_\mathrm{d}, \phi)$ represent the elevation and azimuth-beampatterns, respectively. The quantity, $|\mathcal{B}(f, \theta, \phi)|^2$, is referred to as the powerpattern, and is generally used for visualization. Thus, $|\mathcal{B}(f, \theta, \phi_\mathrm{d})|^2$ represents the elevation-powerpattern, and $|\mathcal{B}(f, \theta_\mathrm{d}, \phi)|^2$ represents the azimuth-powerpattern. The 3-dB elevation-beamwidth is measured as

$$\begin{aligned} \theta_-(f) &= \underset{\theta:\theta<\theta_\mathrm{d}}{\mathrm{argmax}} \frac{|\mathcal{B}(f, \theta, \phi_\mathrm{d})|}{|\mathcal{B}(f, \theta_\mathrm{d}, \phi_\mathrm{d})|} \leq 0.5, \\ \theta_+(f) &= \underset{\theta:\theta>\theta_\mathrm{d}}{\mathrm{argmin}} \frac{|\mathcal{B}(f, \theta, \phi_\mathrm{d})|}{|\mathcal{B}(f, \theta_\mathrm{d}, \phi_\mathrm{d})|} \leq 0.5, \\ b_\theta(f) &= \theta_+(f) - \theta_-(f). \end{aligned} \tag{8}$$

Similarly, the 3-dB azimuth-beamwidth is measured as

$$\begin{aligned} \phi_-(f) &= \underset{\phi:\phi<\phi_\mathrm{d}}{\mathrm{argmax}} \frac{|\mathcal{B}(f, \theta_\mathrm{d}, \phi)|}{|\mathcal{B}(f, \theta_\mathrm{d}, \phi_\mathrm{d})|} \leq 0.5, \\ \phi_+(f) &= \underset{\phi:\phi>\phi_\mathrm{d}}{\mathrm{argmin}} \frac{|\mathcal{B}(f, \theta_\mathrm{d}, \phi)|}{|\mathcal{B}(f, \theta_\mathrm{d}, \phi_\mathrm{d})|} \leq 0.5, \\ b_\phi(f) &= \phi_+(f) - \phi_-(f). \end{aligned} \tag{9}$$

The principal goal of this work is to obtain an elevation-beamwidth, $b_\theta(f) \geq$ target-elevation-beamwidth (θ_BW), and an azimuth-beamwidth, $b_\phi(f) \geq$ target-azimuth-beamwidth (ϕ_BW), across the speech spectrum. Having

a beamwidth that is not-too-small in both the planes provides immunity against signal distortion if the look-direction does not exactly match the true DOA of the signal-of-interest.

The DF is an important characteristic of the beamformer [1], [2], [22], [23]. It signifies the relative or normalized strength of the beampattern in the look-direction, with respect to the entire 3-D beampattern. It also represents the gain of the beamformer in a diffuse-noise-field. The DF is given by

$$\begin{aligned}\mathcal{D}(f) &= \frac{|\mathcal{B}(f,\theta_{\mathrm{d}},\phi_{\mathrm{d}})|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_{-\pi}^{\pi}|\mathcal{B}(f,\theta,\phi)|^2\sin\theta\ d\phi\ d\theta},\\ &= \frac{|\mathbf{h}^H(f)\mathbf{d}(f,\theta_{\mathrm{d}},\phi_{\mathrm{d}})|^2}{\mathbf{h}^H(f)\Gamma(f)\mathbf{h}(f)},\\ \Gamma(f)|_{i,j} &= \mathrm{sinc}\big[2\pi f l_{i,j}/c\big],\ 1\leq\ i,j\ \leq M.\end{aligned} \tag{10}$$

In the above equation, $\Gamma(f)|_{i,j}$ represents the $(i,j)^{\mathrm{th}}$ position of the matrix, $\Gamma(f)$, and $l_{i,j}$ represents the Euclidean distance between the microphones corresponding to that position. Like the DF, the WNG is an important parameter [1], [2]; the WNG represents the robustness of the beamformer against spatially and temporally uncorrelated noise and is given by

$$\mathcal{W}(f) = \frac{|\mathbf{h}^H(f)\mathbf{d}(f,\theta_{\mathrm{d}},\phi_{\mathrm{d}})|^2}{\mathbf{h}^H(f)\mathbf{h}(f)}. \tag{11}$$

We now proceed towards devising a window-based beamformer that can achieve practical results across the various metrics: elevation-beamwidth, azimuth-beamwidth, DF, and WNG.

III. Window-based beamformer

The quintessential delay-and-sum (DS) beamformer provides appropriate phase compensation to all the microphones and an equal emphasis $(1/M)$ on each of them:

$$\mathbf{h}(f) = \frac{1}{M}\mathbf{d}(f,\theta_{\mathrm{d}},\phi_{\mathrm{d}}). \tag{12}$$

The DS is possibly the simplest beamformer and hence suitable for studying the characteristics of any array. Consider a circular array of arbitrary radius designed with the requisite number of sensors, i.e., a CCA-I with $P=1$ ring and no central microphone. As per (5), as the radius increases/decreases, so does the number of microphones. In the limiting case, when the radius is zero, only one microphone remains. This is the case of the central microphone in a CCA. In general, as the radius of the circular array decreases both the elevation-beamwidth and azimuth-beamwidth (for any particular frequency) increase, and vice versa. An increase in the beamwidths is associated with a decrease in the DF and WNG, and vice versa. The central microphone (zero-radius) possesses no beamforming attributes: it provides an immutable elevation-beamwidth and azimuth-beamwidth of 360^o each, and no directivity and WNG. Hence, the central microphone is avoided in many analytical solutions for frequency-invariant or constant-beamwidth beamforming. Nevertheless, the variation in the characteristics of a circular array with respect to its radius makes a CCA a more versatile array design. By emphasizing or weighting the rings of the CCA differently at different frequencies, one can attempt to obtain uniform beamforming characteristics over a broad range of frequencies.

Weighting the rings of the CCA, however, provides the same weight to all the microphones in any given ring. One could, additionally, weight the individual microphones in each ring differently. That would obviously make the overall weighting process more versatile, and in turn, enable more control on the beamforming characteristics. "How to weight the individual sensors in a ring ?" is an open-ended question, and there could be many possible approaches. In this work, we will use a simple approach based on the alignment of the sensors with the signal-path. To elaborate and analyze further, we must first define our proposed window-based beamformer, which weights the microphones in the CCA not only based on their ring-radius but also their alignment with the signal-path:

$$\begin{aligned}h_{p,m_p}(f) &= w_p(f)\ a_{p,m_p}(f)\ d_{p,m_p}(f,\theta_{\mathrm{d}},\phi_{\mathrm{d}}),\\ h_{p,m_p}(f) &\coloneqq \frac{h_{p,m_p}(f)}{\sum_p\sum_{m_p}|h_{p,m_p}(f)|}\ ;\end{aligned} \tag{13}$$

$$\begin{aligned}&w_p(f)\in[0,\ 1],\ \sum_p w_p(f)=1,\ \text{and}\\ &a_{p,m_p}(f) = \begin{cases}1,\ p=0\\ \frac{I_0\left(\beta_p(f)\sqrt{1-v_{p,m_p}^2}\right)}{I_0(\beta_p(f))},\ p\geq 1\end{cases},\ \text{where}\end{aligned} \tag{14}$$

$$\begin{aligned}v_{p,m_p} &= \begin{cases}||\mathbf{u}_{\theta_{\mathrm{d}},\phi_{\mathrm{d}}}-\mathbf{u}_{\frac{\pi}{2},\phi_{p,m_p}}||_2,\ \phi_{p,m_p}\in[\phi_{\mathrm{d}}-\frac{\pi}{2},\phi_{\mathrm{d}}+\frac{\pi}{2}]\\ ||\mathbf{u}_{\theta_{\mathrm{d}}+\pi,\phi_{\mathrm{d}}+\pi}-\mathbf{u}_{\frac{\pi}{2},\phi_{p,m_p}}||_2,\ \text{otherwise}\end{cases},\\ v_{p,m_p} &\coloneqq (v_{p,m_p}-v_{\min})/(v_{\max}-v_{\min}),\\ v_{\min} &= ||\mathbf{u}_{\theta_{\mathrm{d}},\phi_{\mathrm{d}}}-\mathbf{u}_{\frac{\pi}{2},\phi_{\mathrm{d}}}||_2,\\ v_{\max} &= ||\mathbf{u}_{\theta_{\mathrm{d}},\phi_{\mathrm{d}}}-\mathbf{u}_{\frac{\pi}{2},\phi_{\mathrm{d}}+\frac{\pi}{2}}||_2.\end{aligned} \tag{15}$$

In (13), $w_p(f)$ denotes the ring-weight, i.e., the weight assigned to the microphone based on its ring-number (hence radius). $a_{p,m_p}(f)$ is the Kaiser-window function that controls the weight assigned to the microphone based on its alignment with the signal-path or line-of-sight of the impinging signal-of-interest. In (14), $\beta_p(f)$ controls the width of the Kaiser-window. $I_0(.)$ denotes the 0^{th} order Bessel function of the first kind. The distance, v_{p,m_p}, between the unit vectors of the microphone-position and the signal-path is used as the input to the Kaiser-window function. In (15), $\mathbf{u}_{\theta,\phi}$ represents the unit vector in Cartesian co-ordinates corresponding to the direction (θ,ϕ) in spherical co-ordinates.

Assume that the DOA is $(45^o,45^o)$ and a particular microphone in a ring is located at $(90^o,30^o)$. This microphone is misaligned with the DOA by $(45^o,15^o)$. The diametrically opposite microphone, located at $(90^o,-150^o)$, will also be misaligned with the signal-path by $(45^o,15^o)$. Both of these microphones must be equally weighted as per their

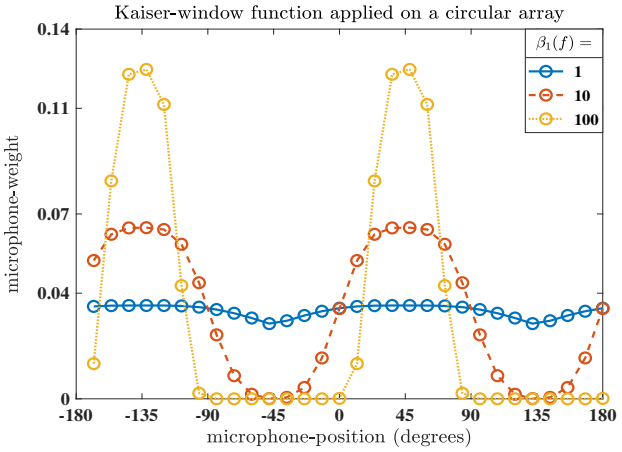

Fig. 3: **Weights assigned to the 30 microphones of a circular array (10 cm radius) for three different values of the Kaiser-window width parameter of the window-based beamformer. The DOA is oblique ($\theta_{\rm d} = \phi_{\rm d} = 45^o$).**

alignment with the DOA. For this purpose, apart from the actual signal-of-interest with DOA, $(\theta_{\rm d}, \phi_{\rm d})$, an imaginary signal-of-interest arriving from $(\theta_{\rm d} + \pi, \phi_{\rm d} + \pi)$ is used in (15). This results in two identical Kaiser-windows for two halves of the ring. Also, note that since the microphones are fixed to the horizontal plane whereas the DOA is arbitrary, a microphone is best aligned when its angular position matches the azimuth-look-direction. In this case, its unit vector, $\mathbf{u}_{\frac{\pi}{2},\phi_{\rm d}}$, will have a minimum distance, $v_{\rm min}$, with the look-direction unit vector, $\mathbf{u}_{\theta_{\rm d},\phi_{\rm d}}$. In the same way, a microphone is completely unaligned when its angular position is orthogonal to the azimuth-look-direction. In this case, its unit vector, $\mathbf{u}_{\frac{\pi}{2},\phi_{\rm d}+\frac{\pi}{2}}$, will have a maximum distance, $v_{\rm max}$, with respect to $\mathbf{u}_{\theta_{\rm d},\phi_{\rm d}}$. Henceforth, $v_{\rm min}$ and $v_{\rm max}$ are used to normalize the distances between [0,1] prior to feeding them to the Kaiser-window function.

To analyze the impact of differently weighting the microphones in a ring, consider a circular array, i.e, a CCA-I with $P = 1$ ring and no central microphone, having a radius of 10 cm. There are 30 microphones in the circular array, as per equation (5). As there is only one ring, the ring-weight is 1, and the microphone-weight is determined purely by the Kaiser-window function. Consider the general oblique DOA $(\theta_{\rm d} = \phi_{\rm d} = 45^o)$. Fig. 3 plots the microphone weights, $|h_{1,m_1}(f)|\ \forall\ m_1$, for three different values of the Kaiser-window width parameter, $\beta_1(f)$. As is evident, the Kaiser-window width is inversely related to the width parameter, which determines how strictly the microphones in the ring are emphasized based on their alignment with the signal-path. In this case, the microphones whose angular positions are closer to $\phi_{\rm d} = 45^o$ are emphasized more. Their diametrically opposite microphones are weighted identically. In the limiting case, $\beta_1(f) \to \infty$, this results in a two-microphone ULA in the end-fire configuration. At the other extreme, $\beta_1(f) \to 0$, all the microphones are equally weighted, and the final beamformer converges to the DS beamformer.

Fig. 4 plots the metrics for the window-based beamformer as $\beta_1(f)$ varies. The metrics are presented for two different frequencies - 500 Hz and 1 kHz. At very low frequencies ($\lesssim$ 500 Hz), the beamwidths are generally very high ($\sim 360^o$). As the Kaiser-window becomes sharper, i.e., $\beta_1(f)$ increases, more unaligned microphones are de-emphasized. Correspondingly, the beampatterns become narrower, and the DF improves. De-emphasizing the microphones is mathematically equivalent to reducing the number of sensors, and decreases the WNG. When the frequency is not so low (as observed in the 1 kHz case), the azimuth-beamwidth and elevation-beamwidth exhibit a non-linear and inverse relationship with each other. That is why it is important to consider both the elevation-beamwidth and azimuth-beamwidth in any practical algorithm. Nevertheless, at higher frequencies, as more unaligned microphones are de-emphasized, both the DF and WNG decreases. Note that beyond $\beta_1(f) > 100$ the DF and WNG seem to plunge to levels which may render the beamformer practically ineffective. Hence, the Kaiser-window width parameter would be operated within the range, $[0, 100]$, in this work.

As is evident from Fig. 4, incorporating the Kaiser-window function to differently weight the microphones in each ring is likely to provide more versatility in obtaining the desired performance, in addition to differently weighting the rings of the CCA. Consider that the ring-weights and Kaiser-window width parameters are initialized with some pre-defined values. If, for any particular frequency, the elevation-beamwidth and azimuth-beamwidth are wider than their respective targets, $\theta_{\rm BW}$ and $\phi_{\rm BW}$, then the ring-weights and Kaiser-window width parameters may be tuned to enhance the DF. That should consequently lead to a reduction in the beamwidths. This scenario is generally expected to occur at lower frequencies. If, on the other hand, either the elevation-beamwidth or the azimuth-beamwidth or both are narrower than their targets, then, we may tune the ring-weights and Kaiser-window width parameters so as to widen the corresponding beamwidth(s) to the target value(s). In any of these scenarios, since the overall effect is to weight the microphones, a decent WNG (> 0 dB) is guaranteed, as will be observed in the results section.

We now propose the following modified gradient-descent algorithm to determine the optimal parameters of the proposed beamformer.

Algorithm: Given $\theta_{\rm BW}, \phi_{\rm BW}, \theta_{\rm d}, \phi_{\rm d}$. Consider $\Delta s(f)$ as the step-size, and $\mu(f)$ as the learning-rate of the gradient-descent algorithm. Define:

$$\begin{aligned} \beta_p'(f) &= \log_{100} \beta_p(f) \in (-\infty, 1],\ p = 1, 2, ..., P, \\ \Delta b_\theta'(f) &= \frac{\theta_{\rm BW} - b_\theta(f)}{\pi/2},\ \Delta b_\phi'(f) = \frac{\phi_{\rm BW} - b_\phi(f)}{\pi/2}, \\ \mathcal{D}'(f) &= -\log_{10} \mathcal{D}(f). \end{aligned} \quad (16)$$

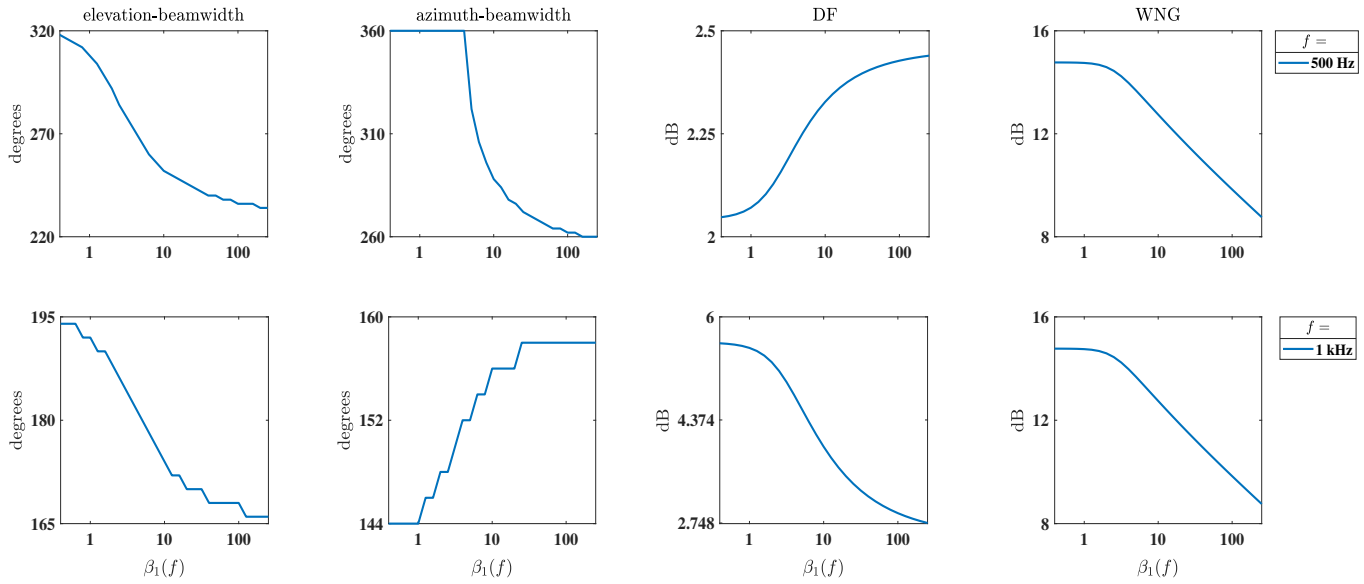

Fig. 4: **Relationship between the performance metrics and the Kaiser-window width parameter of the window-based beamformer, at two frequencies: 500 Hz (top row) and 1 kHz (bottom row). A circular array of 10 cm radius with no central microphone is considered.**

The optimization function is given by:

$$\mathcal{L}(\mathbf{s}(f)) = \begin{cases} \Delta b'_\theta(f),\ \Delta b'_\theta(f) > 0 \text{ and } \Delta b'_\phi(f) \le 0 \\ \Delta b'_\phi(f),\ \Delta b'_\theta(f) \le 0 \text{ and } \Delta b'_\phi(f) > 0 \\ \Delta b'_\theta(f) + \Delta b'_\phi(f),\ \Delta b'_\theta(f) > 0 \text{ and } \Delta b'_\phi(f) > 0 \\ \mathcal{D}'(f),\ \text{otherwise} \end{cases}. \tag{17}$$

In (17), $\mathbf{s}(f)$ represents the set of parameters or independent variables of the optimization function:

$$\begin{aligned} &\mathbf{s}(f) \\ &= [w_0(f),\ ....,\ w_P(f),\ \beta'_1(f),\ ....,\ \beta'_P(f)]^T \\ &= [s_1(f),\ ....,\ s_{P+1}(f),\ s_{P+2}(f),\ ....,\ s_{2P+1}(f)]^T. \end{aligned} \tag{18}$$

Initialize: $\{\mathbf{s}^{(0)}(f) :\ s_l^{(0)}(f) \sim \mathcal{U}(0,\ 1),\ l \in [1,\ 2P+1]\}$, $\Delta s^{(0)}(f) \in (0,1),\ \mu^{(0)}(f) \in (0,1)$.

Iteration: For the n^{th} $(n \ge 0)$ iteration, do the following:

(a) Determine $\mathbf{h}(f)$ using $\mathbf{s}^{(n)}(f)$ and equations (13),(14),(15). Hence, determine $\Delta b'_\theta(f)$ and $\Delta b'_\phi(f)$, and select the optimization function, $\mathcal{L}(\mathbf{s}(f))$, as per (16),(17). Determine the gradient of the optimization function:

$$\nabla_{\boldsymbol{s}(f)}\mathcal{L}(\mathbf{s}^{(n)}(f)) = [\frac{\partial\mathcal{L}(\mathbf{s}^{(n)}(f))}{s_1(f)}, ..., \frac{\partial\mathcal{L}(\mathbf{s}^{(n)}(f))}{s_{2P+1}(f)}]^T,$$
$$\frac{\partial\mathcal{L}(\mathbf{s}^{(n)}(f))}{s_l(f)} = \frac{\mathcal{L}_l^+(f) - \mathcal{L}_l^-(f)}{\Delta s^{(n)}(f)},$$

where $\mathcal{L}_l^+(f)$ is obtained for

$$\boldsymbol{s}^+(f) = [s_1^{(n)}(f), ..., s_l^+(f), ..., s_{2P+1}^{(n)}(f)]^T,$$
$$s_l^+(f) := \begin{cases} s_l^{(n)}(f) + \Delta s^{(n)}(f)/2,\ l \le P+1 \\ \min[1,\ s_l^{(n)}(f) + \Delta s^{(n)}(f)/2],\ l \ge P+2 \end{cases},$$

and $\mathcal{L}_l^-(f)$ is obtained for

$$\boldsymbol{s}^-(f) = [s_1^{(n)}(f), ..., s_l^-(f), ..., s_{2P+1}^{(n)}(f)]^T,$$
$$s_l^-(f) := \begin{cases} \max[0,\ s_l^{(n)}(f) - \Delta s^{(n)}(f)/2],\ l \le P+1 \\ s_l^{(n)}(f) - \Delta s^{(n)}(f)/2,\ l \ge P+2 \end{cases}.$$

(b) Update:

$$\boldsymbol{s}^{(n+1)}(f) = \boldsymbol{s}^{(n)}(f) - \mu^{(n)}(f)\nabla_{\boldsymbol{s}(f)}\mathcal{L}(\mathbf{s}^{(n)}(f)),$$
$$s_l^{(n+1)}(f) := \max[0,\ s_l^{(n+1)}(f)],\ l \le\ P+1,$$
$$s_l^{(n+1)}(f) := \frac{s_l^{(n+1)}(f)}{\sum_{l=1}^{P+1} s_l^{(n+1)}(f)},\ l \in [1,\ P+1],$$
$$s_l^{(n+1)}(f) := \min[1,\ s_l^{(n+1)}(f)],\ l \ge\ P+2.$$

(c) Determine $\mathbf{h}(f)$ using $\mathbf{s}^{(n+1)}(f)$. Hence, determine $\Delta b'_\theta(f)$, $\Delta b'_\phi(f)$, and $\mathcal{L}(\mathbf{s}(f))$, as in step (a).

(c.1) If $\mathcal{L}(\mathbf{s}(f))$ changes to-or-from $\mathcal{D}'(f)$, then reduce the step-size and learning-rate, and revert back the parameters. Return to step (a).

$$\Delta s^{(n+1)}(f) := \Delta s^{(n)}(f)/2,$$
$$\mu^{(n+1)}(f) := \mu^{(n)}(f)/2,$$
$$s_l^{(n+1)}(f) := s_l^{(n)}(f),\ l \in [1,\ 2P+1].$$

(c.2) If either of the beamwidths changes radically, i.e., if $|\Delta b'_\theta(f) \times \pi/2|$ or $|\Delta b'_\phi(f) \times \pi/2|$ is too-large

(say $\geq 20^o$), reduce the step-size and learning-rate. Return to step (a).

$$\Delta s^{(n+1)}(f) \coloneqq \Delta s^{(n)}(f)/2,$$
$$\mu^{(n+1)}(f) \coloneqq \mu^{(n)}(f)/2.$$

(c.3) If either of the beamwidths are slow-to-change, i.e., if $\Delta b'_\theta(f) \times \pi/2$ or $\Delta b'_\phi(f) \times \pi/2$ are varying minutely (say $\leq 2^o$) over two or three consecutive iterations, increase the step-size and learning-rate. Return to step (a).

$$\Delta s^{(n+1)}(f) \coloneqq 2\ \Delta s^{(n)}(f),$$
$$\mu^{(n+1)}(f) \coloneqq 2\ \mu^{(n)}(f).$$

(d) Terminate the algorithm after a pre-defined maximum number of iterations, or when the optimization function, $\mathcal{L}(\mathbf{s}(f))$, saturates.

We denote the resulting beamformer as the Kaiser-window (KW) beamformer. Note that the KW can be applied on any arbitrary CCA, with or without the central microphone. Also, note that unlike in conventional gradient-descent, we utilize 'step-size' and 'learning-rate' as separate parameters. If the mathematical expressions for the gradients, $\{\frac{\partial \mathcal{L}(\mathbf{s}(f))}{s_l(f)},\ l = 1, 2, ..., 2P+1\}$, were available, we could have used the conventional algorithm, where the terms 'step-size' and 'learning-rate' are one and the same. However, in our case, since the mathematical expressions for the gradients are unknown, we use a brute-force gradient-descent algorithm, where the step-size is used to calculate the gradients. A gradient so calculated at a particular point is only an approximation of the true gradient and represents the slope of the line connecting two neighbouring points on either side of the concerned point. We can modify the step-size if the gradient is too-small or too-large to avoid non-change or large fluctuations of the parameters (equivalently that of the optimization function) between iterations. The adaptive learning-rate also serves the same purpose in practice. The MATLAB codes are available online [37].

The convergence of the gradient-descent algorithm is well established in literature [38], [39]. In general, if the learning-rate is small, i.e., $0 < \mu(f) \leq L/2$, where $L > 0$ is the constant of Lipschitz continuity, the gradient-descent algorithm is guaranteed to converge. Computing the value of L is quite expensive and difficult. Thus, in practice, the learning-rate is determined by trial-and-error. Using an adaptive learning-rate facilitates this process. Generally, as we have implemented in our work, a relatively large learning-rate is initially chosen, which is then decreased based on certain checks and balances as the algorithm iterates.

As mentioned earlier, the objective of the KW beamformer and the associated gradient-descent algorithm is to obtain $b_\theta(f) \geq \theta_{\text{BW}}$ and $b_\phi(f) \geq \phi_{\text{BW}}$, while maximizing the DF whenever and wherever possible in the entire frequency spectrum. Recall that unlike known frequency-invariant or constant-beamwidth beamforming methods, the goal is not to attain a fixed beamwidth, but rather beamwidth greater than a lower threshold (and in both the planes, not just the azimuth). To understand the algorithm, consider a CCA-I design, with $P = 4$ rings and a central microphone. The rings are of radii: 5 cm, 10 cm, 15 cm, and 20 cm. Consider the general oblique look-direction ($\theta_{\text{d}} = \phi_{\text{d}} = 45^o$), and the target beamwidths, $\theta_{\text{BW}} = \phi_{\text{BW}} = 40^o$. We use $s_l^{(0)}(f) = 1/(P+1),\ l \in [1,\ P+1]$, and $s_l^{(0)}(f) = 0.5,\ l \in [P+2,\ 2P+1]$. Also, we use $\Delta s^{(0)}(f) = 0.2$ and $\mu^{(0)}(f) = 0.1$ in the algorithm.

Fig. 5 presents the metrics as the number of iterations increases, for two frequencies: 1 kHz and 6 kHz. At 1 kHz, the elevation-beamwidth and azimuth-beamwidth are initially wider than their target (40^o), and hence the DF is maximized. Consequently, the azimuth-beamwidth decreases, whereas the elevation-beamwidth remains almost unchanged. The increase in the DF is associated with a decrease in the WNG, which suggests that unaligned microphones in certain rings were de-emphasized. At 6 kHz, both the elevation-beamwidth and azimuth-beamwidth are initially narrower than their targets, and hence they are widened till they reach their target. When both the elevation and azimuth beamwidth targets are reached, the algorithm switches between maximizing directivity vs. beamwidth control in the subsequent iterations. The checks and balances along with the adaptive step-size and learning-rate ensure that the algorithm saturates without large fluctuations. Note that if there was only a single ring (circular array), increasing the azimuth-beamwidth would likely have decreased the elevation-beamwidth, and vice-versa, as observed in Fig. 4. The presence of multiple rings and the central microphone in the CCA, however, allows much more control on the beamformer characteristics. The increase in the elevation and azimuth beamwidths is associated with the decrease in the DF and WNG, which is expected at high frequency, as discussed earlier.

Fig. 6 shows the optimal parameters of the KW beamformer for the entire frequency spectrum. There are five ring-weight parameters - one for the central microphone ($p = 0$), and the rest for the four rings. Each of the four rings, additionally, have a Kaiser-window width parameter. As we have mentioned in Sec. III, higher frequencies generally have smaller beamwidths, and larger rings further accentuate the situation. Hence, the algorithm assigns larger weights to the central sensor and the smaller rings, as the frequency increases. Contrarily, the algorithm assigns the largest weight to the outermost ring ($p = 4$) at low frequencies. In the case of the Kaiser-window width parameters, it is difficult to intuitively make sense of the optimal parameters as they have a complicated relationship with the metrics. One might simply view them as lending more control over the characteristics of the beamformer, in conjunction with the ring-weight parameters.

IV. Results

In this section, we compare the performance of the KW beamformer with other beamformers. For this purpose, we

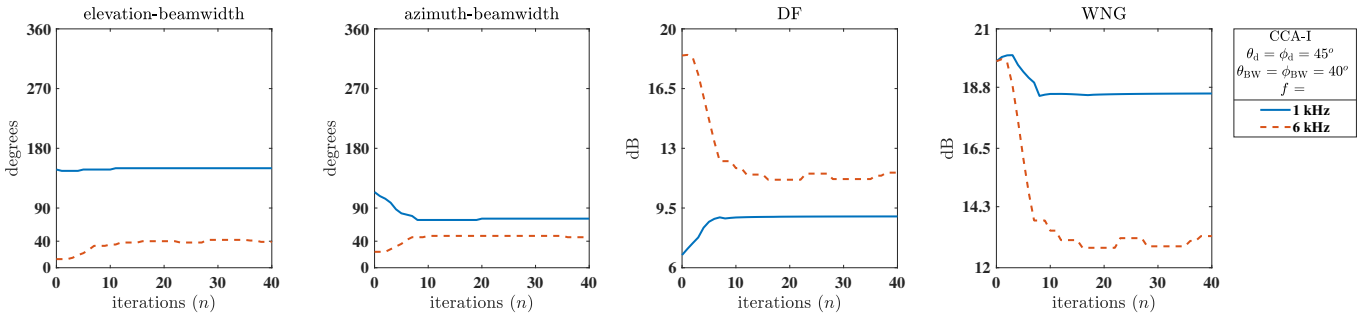

Fig. 5: **Performance metrics for increasing iterations of the gradient-descent-algorithm used to obtain the KW beamformer. The plots are shown for two different frequencies: 1 kHz and 6 kHz. A CCA-I with $P = 4$ rings and a central microphone is considered.**

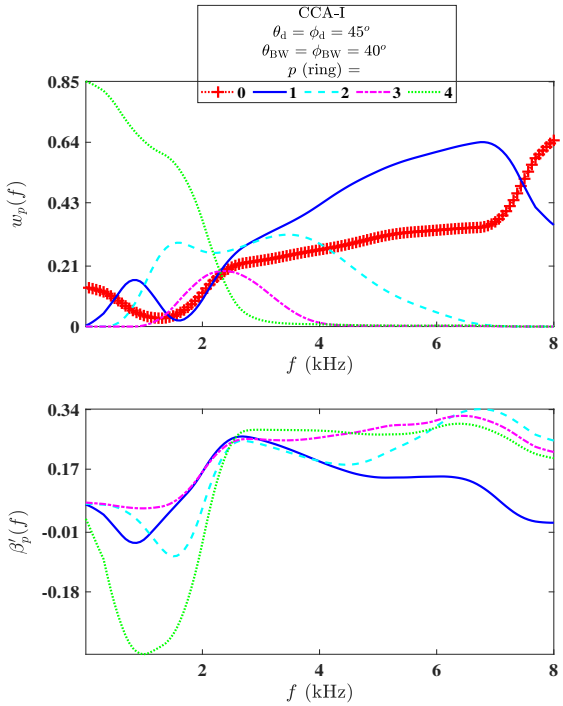

Fig. 6: **Final optimized parameters vs. frequency for the KW beamformer. A CCA-I with $P = 4$ rings and a central microphone is considered.**

consider the same CCA-I with $P = 4$ rings and a central microphone, as used in the previous section. We also consider the CCA-II for the same ring-configuration, which has fewer microphones and a larger inter-sensor distance. We consider the general oblique look-direction ($\theta_d = \phi_d = 45^o$), and the target-beamwidths, $\theta_{BW} = \phi_{BW} = 40^o$. There is no particular reason as to why we have chosen the target beamwidths as 40^o. The KW beamformer is independent of the choice of the target beamwidths. In general, the higher the target beamwidths worse will be the DF and WNG. The suitable targets, hence, should be chosen based on the overall performance requirements across all the metrics. In the proposed modified gradient-descent algorithm, the same algorithmic-parameters are used. The KW beamformer is compared with two other state-of-the-art beamformers, briefly discussed below:

Method-1 [18]: The DS beamformer is obtained for each ring of the CCA, i.e., $\mathbf{h}_p(f) = \mathbf{d}_p(f, \theta_d, \phi_d)/M$. For each ring, the frequency at which the target-azimuth-beamwidth is best met is considered the reference frequency, and the corresponding DS beamformer is the reference beamformer. For each ring, the beamformer at any given frequency is determined analytically using its reference beamformer and a transformation matrix derived for the particular frequency using the Bessel-function approximation. The beamformer for the central microphone is kept the same as that of the DS beamformer, i.e., $\mathbf{h}_0(f) = h_{0,0}(f) = 1/M$. The amplitudes of the final M-dimensional beamformer are normalized.

Method-2 [22], [23]: For the CCA, a spatial-filter is designed which provides the desired azimuth-beampattern irrespective of frequency. In this work, we use a symmetric hamming-window FIR filter (12-order) which meets the target-azimuth-beamwidth requirement. Beamformers at each frequency are then designed based on the optimal approximation of the beampattern with the Jacobi-Anger expansion.

Additionally, the performance of the DS beamformer is also included for completeness of the study. Note that solutions for frequency-invariant or constant-beamwidth beamforming for planar arrays are focussed on controlling the azimuth-beampattern only. The same applies to Method-1 and Method-2. The choice of Method-1 and Method-2, in this context, is purely because they are able to perfectly attain the target-azimuth-beamwidth across the entire frequency spectrum, as we will observe shortly.

A. Performance for CCA-I

Consider the CCA-I design. As is evident from Fig. 7, Method-1 and Method-2 are able to perfectly attain the target, $\phi_{BW} = 40^o$, across the entire frequency spectrum. However, they provide no control over the elevation-beampattern. In the low-frequency spectrum ($\lesssim 2.5$ kHz), the WNG is $\lesssim 0$ dB, and hence the solution is arguably impractical. As we have mentioned in the introduction section, lower frequencies have naturally wider beamwidths, which when forced to be much narrower, might be detrimental. As the frequency increases, the WNG improves for both Method-1 and Method-2. In fact, at higher frequencies, both the DF and WNG of Method-2 are better than that of the KW. However, this

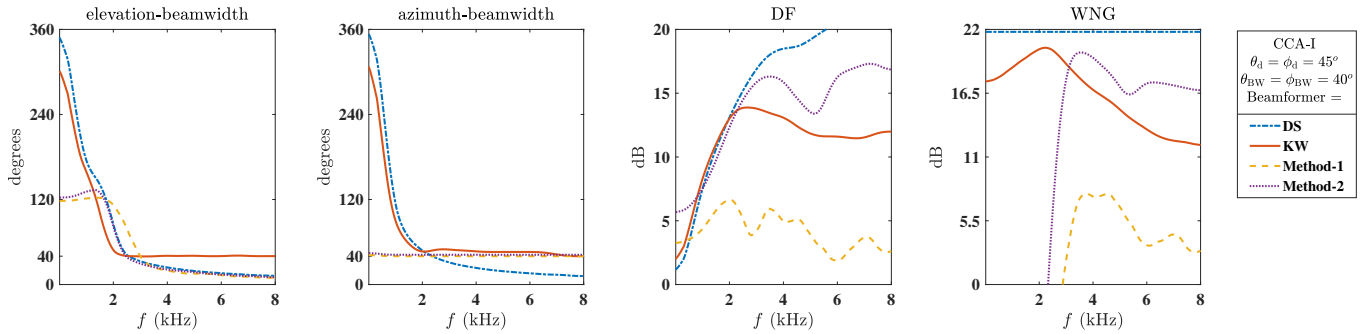

Fig. 7: **Comparison of the performance metrics of the DS, KW, Method-1, and Method-2 beamformers applied on a CCA-I with $P = 4$ rings and a central microphone.**

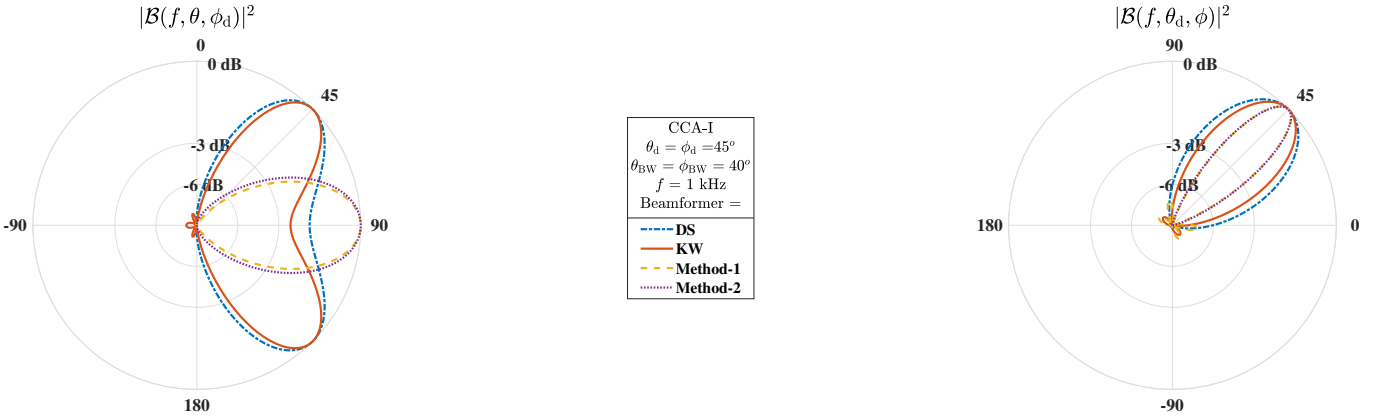

Fig. 8: **Comparison of the elevation and azimuth powerpatterns (normalized) at 1 kHz frequency for the DS, KW, Method-1, and Method-2 beamformers. A CCA-I of $P = 4$ rings and a central microphone is considered.**

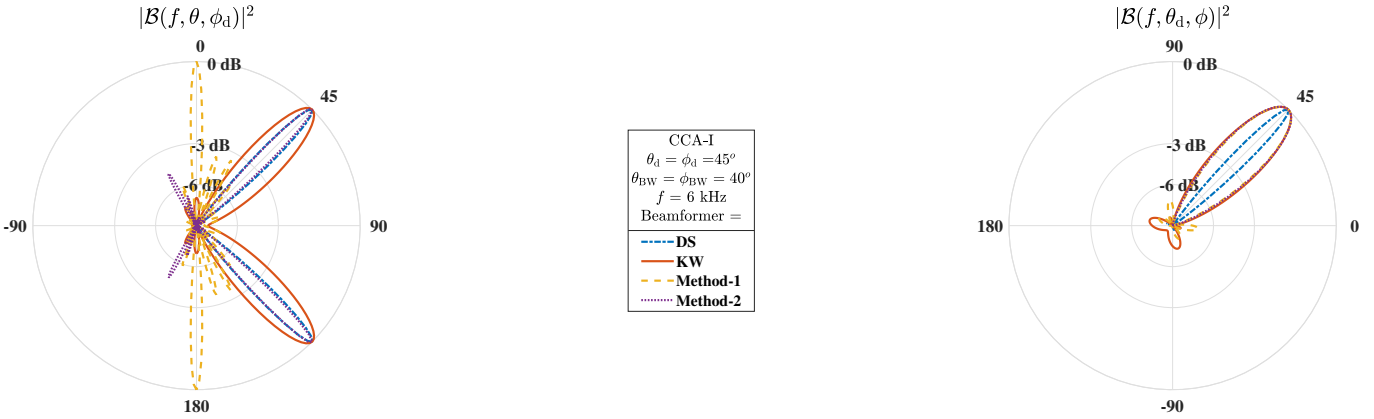

Fig. 9: **Comparison of the elevation and azimuth powerpatterns (normalized) at 6 kHz frequency for the DS, KW, Method-1, and Method-2 beamformers. A CCA-I of $P = 4$ rings and a central microphone is considered.**

is because the elevation-beamwidth is narrower than the desired target, $\theta_{\text{BW}} = 40^o$. The same phenomenon occurs in the case of the DS beamformer. For the DS, at high frequencies, both the elevation and azimuth beamwidths are narrower than their targets; hence it has the highest DF. Notice, however, that at low frequencies, the KW provides narrower beamwidths and better DF than the DS. The relatively poor DF and WNG of Method-1 is due to the misdirection phenomenon, to be discussed next.

Fig. 8 shows the elevation and azimuth powerpatterns for the four methods at 1 kHz. As can be observed, unlike the KW and the DS, the other two methods have the perfect azimuth-powerpatterns pointing towards $\phi_{\text{d}} = 45^o$. However, their elevation-powerpatterns are misdirected towards the horizontal plane, instead of $\theta_{\text{d}} = 45^o$. In other words, the main beam is not pointing towards the DOA. This phenomenon is observed in the entire low-frequency spectrum, for the two analytical methods. As is obvious, if $\theta_{\text{d}} = 90^o$, the misdirection phenomenon would not be witnessed. Consequently, the WNG would be much higher (> 0 dB), as reported in the relevant works. This exemplifies why analytical solutions are confined to the horizontal plane, as we have mentioned in the introduction section. Nevertheless, the beamwidths

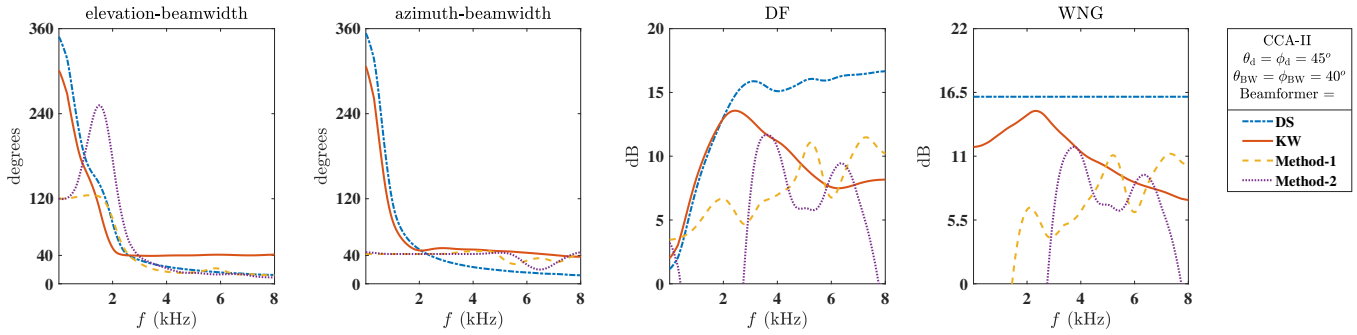

Fig. 10: **Comparison of the performance metrics of the DS, KW, Method-1, and Method-2 beamformers applied on a CCA-II with $P = 4$ rings and a central microphone.**

calculated for the two methods, using equations (8) and (9), become irrelevant when $\theta_d \approx 90^o$. The KW and DS beamformers, contrary to the other two methods, do not exhibit this undesirable phenomenon. Further, the KW has narrower beamwidths than the DS, striving to provide a practical, though not perfect, solution in the low-frequency spectrum.

Fig. 9 shows the elevation and azimuth powerpatterns for the four methods at 6 kHz. As is clear from the figure, the misdirection phenomenon does not persist at higher frequencies for Method-2. However, the elevation-powerpattern is narrower than required. For the DS, both the elevation and azimuth powerpatterns are narrower than required. Only the KW is able to obtain the desired powerpatterns. For Method-1, the misdirection phenomenon continues to persist even at higher frequencies but aimed at random directions and not just the horizontal plane. This explains why its DF and WNG are worse than that of the other methods, as observed in Fig. 7.

The above observations show the significance of the elevation-beampattern and its effect on the performance metrics. These observations, in both the low-frequency and the high-frequency spectrum, are of utmost significance for speech signals. It is well-established that voiced speech (generally vowels) is characterized by its low-frequency spectrum, which contains the pitch or fundamental frequency of the speaker and the principal vocal tract resonances or formants; whereas unvoiced speech (generally consonants) is characterized by its high-frequency spectrum [14], [15]. As such, the practical utility and advantage of the proposed KW beamformer are obvious with respect to known solutions.

Here, one may wonder if the performance of the KW beamformer would remain the same if the central microphone is removed from the CCA-I design. That is indeed the case. In fact, the performances of all the four methods are observed to be nearly identical with and without the central microphone. As such, we have not included the corresponding figures for the CCA-I design without the central microphone. However, had there been only a couple of rings in the CCA-I design (in the extreme case, it is a circular array of radius 20 cm), then, the central microphone could have played a significant role in widening the beamwidths at high frequencies. In general, when multiple rings of large and small sizes are present in the CCA design, as in the CCA-I, the central microphone is not so influential. Nevertheless, the proposed KW beamformer (and the associated gradient-descent algorithm) is design-independent: it has no specific design requirements.

B. Performance for CCA-II

Fig. 10 plots the metrics for the 4-ring CCA-II design, where the number of microphones in each ring is a quarter of that in the CCA-I design. The KW and the DS continue to behave similarly to the case of the CCA-I design. Their DFs and WNGs are expectedly worse since the number of microphones is decreased. The other two beamformers, however, exhibit marked misbehavior. As is evident, at higher frequencies, Method-1 and Method-2 are unable to meet the target, $\phi_{BW} = 40^o$. For Method-2, the DF and WNG are substantially worse, which is because of the increased severity of the misdirection phenomenon, discussed earlier. This is, however, not unexpected, as Method-2 was specifically designed for smaller CCAs, where $\delta \leq \lambda_{min}/2$ is always maintained. Equivalent requirements of substantially large or a certain specific number of microphones also exist in other methods, which indirectly keeps the inter-sensor distance in check [16], [17], [33], [40]. Hence, as we have mentioned earlier, in many solutions the central microphone is not included in the CCA design. Method-1, it seems, is more robust to variations in the design of the CCA. This is likely because its solution is independent of any requirement regarding the number of microphones. This flexibility, however, comes at the expense of overall poorer performance, as we have already witnessed. Its elevation-beampattern (not shown) is, as usual, misdirected.

C. Speech example: outputs for misestimated DOA

We consider the 4-ring CCA-II design (for no particular reason). A speech signal-of-interest (arbitrarily taken from the TIMIT corpus [41]), impinges on the array with the DOA, $(60^o, 60^o)$. A high-frequency channel interference (obtained from the NOISEX-92 corpus [42]) also impinges

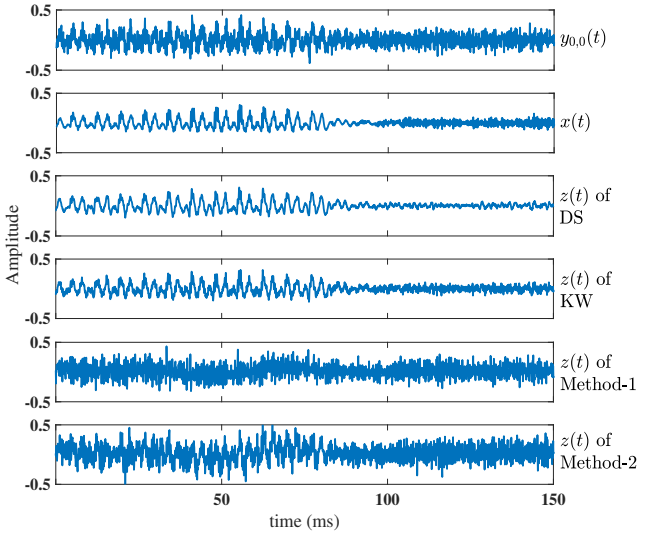

Fig. 11: **Comparison of the time-domain outputs of the DS, KW, Method-1, and Method-2 beamformers applied on a CCA-II with $P = 4$ rings and a central microphone. $y_{0,0}(t)$ is the combined signal received at the central microphone, $x(t)$ is the impinging speech signal, and $z(t)$ represents the output after beamforming. iSIR = 0 dB, iSNR = 10 dB.**

on the array at an angle $(90^o, 90^o)$. The input signal-to-interference ratio (iSIR) is 0 dB. All the microphones in the array are afflicted with their own thermal-noise, with the input signal-to-noise ratio (iSNR) being 10 dB. The DOA of the speech signal is misestimated and the look-direction is considered as $(45^o, 45^o)$. Based on this look-direction, the four beamformers are derived and then applied to the overall data impinging on the array.

The first (topmost) plot of Fig. 11 shows the noisy data received by the central microphone. The second plot shows the pure speech signal, and the last four plots show the final outputs of the beamformers. As is evident, the KW beamformer output is the closest match to the speech signal. The DS produces a cleaner output but causes loss of high-frequency information, particularly in the unvoiced region beyond 100 ms. The DS behaves like a low-pass filter in this context. Kindly note that metrics such as the output SNR, PESQ, and STOI may not reveal such important differences (may be investigated in the future), as they provide an average score for the entire speech signal. The other two methods, owing to their limitations discussed earlier, do not provide comparable outputs.

V. Summary and Conclusions

This work presents a broadband beamforming methodology for any arbitrary CCA of microphones and any arbitrary look-direction. The objective is to control both the elevation and azimuth beamwidths while ensuring a robust DF and WNG, so as to provide a practical beamforming solution across the speech spectrum. The principal idea involves weighting the microphones of the CCA based on their corresponding ring-radii, and their alignment with the line-of-sight of the signal-of-interest. A symmetric-window function, such as Kaiser, Gaussian, Chebyshev, etc., may be used to weight the microphones in any given ring of the CCA. In this work, we have utilized the Kaiser-window. A modified version of the time-trusted gradient-descent algorithm is utilized to optimize the ring-weights and the Kaiser-window width parameters, both of which determine the final weights of the microphones. The resulting beamformer is called the KW beamformer. The results show the flexibility and advantage of the proposed methodology over state-of-the-art methods:

- The KW beamformer enables control of not only the azimuth-beamwidth, but also the elevation-beamwidth.
- Whether the CCA contains the requisite number of microphones to avoid spatial aliasing or the CCA design is sparse and has a limited number of microphones, the operation of the KW beamformer is uninfluenced, though the final performance may vary. Likewise, the KW beamformer has no inherent limitations pertaining to the central microphone.
- The KW beamformer is designed for any arbitrary DOA of the signal-of-interest, more crucially the elevation angle of the DOA. It is not limited to signals in the horizontal plane.

As we have discussed in the introduction and the results, it is difficult to analytically construct a beamformer which provides control on multiple facets. Only under certain conditions, such methods may result in optimal performance across all metrics. Our proposed approach converts the problem statement to a numerical framework, where gradient-descent can be intelligently utilized to meet our goals. In this work, we have demonstrated how the optimization function, $\mathcal{L}(\mathbf{s}(f))$, can be switched between beamwidth and directivity, as necessary. Of course, it is also possible to define $\mathcal{L}(\mathbf{s}(f))$ as a weighted sum of multiple metrics, which may be experimented with in the future. Similarly, experimentation with different symmetric-window functions may be carried out. Other approaches to weight the sensors in the rings could also be researched. The drawback of our methodology is its computational-time: while we have implemented our work in MATLAB, and other programming domains would be much faster, gradient-descent would likely be slower than any analytical solution. However, by pre-evaluating and storing the microphone-weights for a range of look-directions, we may utilize a sub-optimal solution based on the closest match of the required look-direction.

References

[1] J. Benesty, I. Cohen, and J. Chen, *Fundamentals of Signal Enhancement and Array Signal Processing*. John Wiley & Sons, 2018.

[2] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*. Springer Science & Business Media, 2008, vol. 1.

[3] X. Wang, I. Cohen, J. Chen, and J. Benesty, "On robust and high directive beamforming with small-spacing microphone arrays for scattered sources," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 27, no. 4, p. 842–852, Apr. 2019. [Online]. Available: https://doi.org/10.1109/TASLP.2019.2899517

[4] C. Pan, J. Chen, J. Benesty, and G. Shi, "On the design of target beampatterns for differential microphone arrays," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 27, no. 8, p. 1295–1307, Aug. 2019. [Online]. Available: https://doi.org/10.1109/TASLP.2019.2918081

[5] W. Li, L. Wang, Y. Zhou, J. Dines, M. Magimai.-Doss, H. Bourlard, and Q. Liao, "Feature mapping of multiple beamformed sources for robust overlapping speech recognition using a microphone array," *IEEE/ACM Trans. Audio, Speech and Lang. Proc.*, vol. 22, no. 12, p. 2244–2255, Dec. 2014. [Online]. Available: https://doi.org/10.1109/TASLP.2014.2364130

[6] K. Kumatani, J. McDonough, and B. Raj, "Microphone array processing for distant speech recognition: From close-talking microphones to far-field sensors," *IEEE Signal Processing Magazine*, vol. 29, no. 6, pp. 127–140, 2012.

[7] M. J. Taghizadeh, P. N. Garner, and H. Bourlard, "Microphone array beampattern characterization for hands-free speech applications," in *2012 IEEE 7th Sensor Array and Multichannel Signal Processing Workshop (SAM)*. IEEE, 2012, pp. 465–468.

[8] R. Sharma, I. Cohen, and J. Benesty, "Adaptive and hybrid kronecker product beamforming for far-field speech signals," *Speech Communication*, vol. 120, pp. 42–52, 2020.

[9] 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*, 2020.

[10] G. Huang, J. Chen, J. Benesty, I. Cohen, and X. Zhao, "Steerable differential beamformers with planar microphone arrays," *EURASIP Journal on Audio, Speech, and Music Processing*, vol. 2020, no. 1, pp. 1–18, 2020.

[11] G. Huang, I. Cohen, J. Chen, and J. Benesty, "Continuously steerable differential beamformers with null constraints for circular microphone arrays," *The Journal of the Acoustical Society of America*, vol. 148, no. 3, pp. 1248–1258, 2020.

[12] J. Benesty, J. Chen, and I. Cohen, "Design of third-order circular differential arrays," in *Design of Circular Differential Microphone Arrays*. Springer, 2015, pp. 81–90.

[13] 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.

[14] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer handbook of speech processing*. Springer Science & Business Media, 2008.

[15] R. Sharma, L. Vignolo, G. Schlotthauer, M. Colominas, H. L. Rufiner, and S. Prasanna, "Empirical mode decomposition for adaptive AM-FM analysis of speech: A review," *Speech Communication*, vol. 88, pp. 39 – 64, 2017. [Online]. Available: //www.sciencedirect.com/science/article/pii/S0167639316302370

[16] S. Chan and 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, 2006.

[17] S.-C. Chan and H. Chen, "Theory and design of uniform concentric circular arrays with frequency invariant characteristics [sensor arrays]," in *Proceedings.(ICASSP'05). IEEE International Conference on Acoustics, Speech, and Signal Processing, 2005.*, vol. 4. IEEE, 2005, pp. iv–805.

[18] Y. Yang, C. Sun, and C. Wan, "Theoretical and experimental studies on broadband constant beamwidth beamforming for circular arrays," in *Oceans 2003. Celebrating the Past... Teaming Toward the Future (IEEE Cat. No. 03CH37492)*, vol. 3. IEEE, 2003, pp. 1647–1653.

[19] H. Steyskal, "Circular array with frequency-invariant pattern," in *Digest on Antennas and Propagation Society International Symposium*. IEEE, 1989, pp. 1477–1480.

[20] H. Chen, S. Chan, and K. Ho, "Adaptive beamforming using frequency invariant uniform concentric circular arrays," *IEEE Transactions on Circuits and Systems I: Regular Papers*, vol. 54, no. 9, pp. 1938–1949, 2007.

[21] W. Liu and S. Weiss, "Design of frequency invariant beamformers for broadband arrays," *IEEE Transactions on Signal Processing*, vol. 56, no. 2, pp. 855–860, 2008.

[22] G. Huang, J. Chen, and J. Benesty, "On the design of robust steerable frequency-invariant beampatterns with concentric circular microphone arrays," in *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2018, pp. 506–510.

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

[24] Y. Buchris, A. Amar, J. Benesty, and I. Cohen, "Incoherent synthesis of sparse arrays for frequency-invariant beamforming," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 3, pp. 482–495, 2018.

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

[26] T. Long, I. Cohen, B. Berdugo, Y. Yang, and J. Chen, "Window-based constant beamwidth beamformer," *Sensors*, vol. 19, no. 9, p. 2091, 2019.

[27] P. Ioannides and C. A. Balanis, "Uniform circular and rectangular arrays for adaptive beamforming applications," *IEEE Antennas and Wireless Propagation Letters*, vol. 4, pp. 351–354, 2005.

[28] J. Davis and A. Gibson, "Phase mode excitation in beamforming arrays," in *2006 European Radar Conference*. IEEE, 2006, pp. 307–310.

[29] M. Askari and M. Karimi, "Sector beamforming with uniform circular array antennas using phase mode transformation," in *2013 21st Iranian Conference on Electrical Engineering (ICEE)*. IEEE, 2013, pp. 1–6.

[30] H. Pessentheiner, S. Petrik, and H. Romsdorfer, "Beamforming using uniform circular arrays for distant speech recognition in reverberant environments and double talk scenarios," in *Thirteenth Annual Conference of the International Speech Communication Association*, 2012.

[31] M. Reza, M. Hossain, M. Hossein, and M. Rashid, "Robust uniform concentric circular array beamforming in the existence of look direction disparity," in *2016 2nd International Conference on Electrical, Computer & Telecommunication Engineering (ICECTE)*. IEEE, 2016, pp. 1–4.

[32] Y. Buchris, I. Cohen, J. Benesty, and A. Amar, "Joint sparse concentric array design for frequency and rotationally invariant beampattern," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1143–1158, 2020.

[33] A. Kleiman, I. Cohen, and B. Berdugo, "Insights into fixed constant-beamwidth beamforming with concentric circular arrays," *submitted to IEEE Transactions on Signal Processing*, 2020.

[34] M. I. Dessouky, H. A. Sharshar, and Y. A. Albagory, "Efficient sidelobe reduction technique for small-sized concentric circular arrays," *Progress In Electromagnetics Research*, vol. 65, pp. 187–200, 2006.

[35] M. Dessouky, H. Sharshar, and Y. Albagory, "A novel tapered beamforming window for uniform concentric circular arrays," *Journal of Electromagnetic Waves and Applications*, vol. 20, no. 14, pp. 2077–2089, 2006.

[36] A. Dalli, L. Zenkouar, E. Adiba, M. Habibi, and S. Bri, "Circular array with central element for smart antenna," *Electrical and Electronic Engineering*, vol. 3, no. 3, pp. 86–95, 2013.

[37] [Online]. Available: https://github.com/rajibd2k/KW-beamformer.git

[38] [Online]. Available: https://www.stat.cmu.edu/~ryantibs/convexopt-F13/scribes/lec6.pdf

[39] [Online]. Available: https://www.cs.ubc.ca/~schmidtm/Courses/540-W18/L4.pdf

[40] S.-C. Chan and C. K. Pun, "On the design of digital broadband beamformer for uniform circular array with frequency invariant characteristics," in *2002 IEEE International Symposium on Circuits and Systems. Proceedings (Cat. No. 02CH37353)*, vol. 1. IEEE, 2002, pp. I–I.

[41] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, and N. L. Dahlgren, "DARPA TIMIT acoustic phonetic continuous speech corpus CDROM," 1993. [Online]. Available: http://www.ldc.upenn.edu/Catalog/LDC93S1.html

[42] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: Ii. noisex-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech communication*, vol. 12, no. 3, pp. 247–251, 1993.