Speech Communication 149 (2023) 98–107

Contents lists available at ScienceDirect

Speech Communication

journal homepage: www.elsevier.com/locate/specom

Differential constant-beamwidth beamforming with cube arrays ☆

Gal Itzhak *, Israel Cohen

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

ARTICLE INFO

Keywords:
Microphone arrays
Constant beamwidth beamformer
Differential beamforming
Cube arrays
Three-dimensional arrays

ABSTRACT

In this paper, we present an azimuth and elevation constant-beamwidth (CB) differential beamforming approach for cube arrays. We decompose a global cube beamformer into a Kronecker-product (KP) of three sub-beamformers: two constant-beamwidth beamformers along the y and z axes, and a tunable super-directive (SD) beamformer along the x-axis. We propose two design methods to derive cube beamformers whose either white noise gain (WNG) or directivity factor (DF) may be set by design. We show that the CB threshold frequency with respect to the azimuth and elevation angles, and the WNG and DF performance, can be controlled by the number of microphones along each axis. In addition, we focus on the particular case of merely a single microphone along the x-axis, which yields a rectangular azimuth and elevation CB beamformer. We demonstrate that its CB threshold frequencies are controlled by the number of microphones along the y and z axes, and show that the WNG and DF performance measures are maximized when the two axes are equal in size. Finally, we analyze the performance of the proposed beamformers through simulations of speech signals in various reverberant scenarios, including deviations in the desired speech signal's direction of arrival (DOA). We show that the proposed beamformers outperform previously-presented beamformers and the traditional SD and delay-and-sum (DS) beamformers, particularly in terms of the intelligibility of their corresponding time-domain enhanced signals.

1. Introduction

In the past few decades, beamforming design has been a fruitful research topic yielding a variety of approaches for numerous possible applications. The primary objective is to estimate signals of interest out of noisy observations simultaneously sampled in different locations in space (Benesty et al., 2018; Johnson and Dudgeon, 1992; Van Trees, 2004). Directly applied to the noisy observations, time-domain beamformers are the easiest to implement, yet they typically generate a single speech sample estimate at a time. Indeed, it is possible to estimate a vector of successive speech samples simultaneously. However, such beamformers tend to suffer from high computational complexity (Benesty et al., 2018; Buchris et al., 2019).

Frequency-domain beamformers, as in Buchris et al. (2018), Itzhak et al. (2019), Jin et al. (2021) and Itzhak et al. (2021), are typically formulated on a frame basis, implying that the noisy observations are first transformed into the frequency (or time–frequency) domain by invoking the short-time Fourier transform (STFT), processed, and then transformed back to the time domain to obtain an enhanced version of the sampled noisy observations. Such beamformers generate a frame of enhanced time-domain samples at a time, making them more efficient than time-domain beamformers and hence more appealing for practical applications.

Typically, beamforming is designed and applied to arrays whose geometries are relatively simple, such as uniform linear arrays (ULAs), as their properties are the easiest to control and analyze. Nevertheless, their simple nature results in inherent drawbacks. For example, the phase difference between every two adjacent microphones is identical. This implies that a ULA may only sense the desired signal from a single perspective.

More sophisticated array geometries have been proposed to enhance the beamforming sensing perspective in either two-dimensional (2-D) (Benesty et al., 2015; Huang et al., 2018) or three-dimensional (3-D) (Rafaely, 2015; Wang et al., 2021) layouts. In particular, uniform rectangular arrays (URAs) have been shown valuable for direction of arrival (DOA) estimation methods (Zoltowski et al., 1996; Heidenreich et al., 2012), and in the context of differential beamforming, that is, when the interelement spacing (along both axes) is small (Itzhak et al., 2021, 2022). In contrast, 3-D geometries are known to allow further spatial sensing capabilities. For example, spherical arrays were shown to enable DOA estimation with respect to both the azimuth and elevation angles (Khaykin and Rafaely, 2009; Moore et al., 2017)

☆ This research was supported by the Pazy Research Foundation.

* Corresponding author.
E-mail addresses: galitz@campus.technion.ac.il (G. Itzhak), icohen@ee.technion.ac.il (I. Cohen).

https://doi.org/10.1016/j.specom.2023.03.006
Received 18 June 2022; Received in revised form 8 December 2022; Accepted 8 March 2023
Available online 27 March 2023
0167-6393/

as well as direct control over a beampattern's main-lobe width and maximum side lobe level (Koretz and Rafaely, 2009).

The concept and application of KP beamforming have significantly evolved in the past few years. It enables flexible design approaches in which a global beamformer is decomposed into a KP of independent sub-beamformers that may be individually designed and optimized (Abramovich et al., 2010; Ribeiro et al., 2016; Benesty et al., 2019; Cohen et al., 2019; Itzhak et al., 2021; Wang et al., 2021). Each sub-beamformer is potentially optimized according to a different criterion, yielding a global beamformer that is "optimized" according to all requirements. The portion of each optimization criterion is typically determined by the relative sizes of the corresponding sub-beamformers.

Broadband applications (e.g., communication and speech signals) suffer from a frequency-varying spatial array response. In practical scenarios, this might lead to distortion of the desired signal as its DOA is not necessarily known in advance with high precision. To address this issue, several approaches have been proposed. One possible approach suggests defining a region of interest (ROI) and optimizing the beamformer considering a continuous range of possible DOAs rather than a single direction, either by array geometry optimization (Konforti et al., 2022) or using a feedback loop to align the desired signal before the beamforming (Zhuang et al., 2019). Alternatively, to avoid undesirable distortion, it is common to employ CB beamformers which maintain a fixed beamwidth over a wide frequency range, typically above a threshold frequency which is a function of the array aperture and the beamforming design technique (Parra, 2006; Tourbabin et al., 2012; Rosen et al., 2017; Long et al., 2019).

We have recently presented a differential, and CB beamforming approach based on KP beamforming with URAs (Itzhak and Cohen, 2022). We have demonstrated the proposed approach to outperform the linear CB beamformer in terms of the array directivity, particularly in high frequencies, and to outperform the linear SD beamformer in terms of robustness to spatially white noise. Nevertheless, this approach may only attain the CB property with respect to a single (e.g., the azimuth) angle.

In this paper, we present an azimuth and elevation CB differential beamforming approach for cube arrays. We decompose a global cube beamformer into a KP of three sub-beamformers: two constant-beamwidth beamformers along the z and y axes, respectively, and an SD beamformer along the x-axis. At some level, this approach may be seen as a 3-D generalization of the approach suggested in Itzhak and Cohen (2022). We propose two design methods to derive global cube beamformers whose either WNG or DF may be set by design. We show that the CB threshold frequency with respect to the azimuth and elevation angles, and the WNG and DF performance, can be controlled by the number of microphones along each axis. In addition, we focus on the particular case of merely a single microphone along the x-axis, which yields a rectangular azimuth and elevation CB beamformer. We demonstrate that its CB threshold frequencies are controlled by the number of microphones along the y and z axes, and show that its WNG and DF measures are maximized when the two axes are equal in size. Finally, we analyze the performance of the proposed beamformers through simulations of speech signals in various reverberant scenarios, including deviations in the desired speech signal's DOA. We show that the proposed beamformers outperform previously-presented beamformers as well as the traditional SD and DS beamformers, particularly in terms of the intelligibility of their corresponding time-domain enhanced signals.

The rest of the paper is organized as follows. In Section 2, we present the signal model and notations used throughout the paper. In Section 3, we briefly discuss the properties of 3-D KP beamforming and derive the performance measures accordingly. Then, in Section 4, we derive our proposed beamformers: two types of cube azimuth and elevation CB differential beamformers and a rectangular azimuth and elevation CB beamformer. Section 5 demonstrates and analyzes the properties of the proposed beamformers through design examples as well as simulations of noisy speech signals in reverberant environments and with deviations of the speech signals' DOA. Finally, we summarize this work in Section 6.

2. Signal model

Let us assume that a signal of interest propagates in the shape of a plane wave from the farfield in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, in an azimuth angle ϕ and an elevation angle θ. The plane wave impinges on a 3-D microphone cube array whose edges define the $\mathrm{x}-\mathrm{y}-\mathrm{z}$ axes of a Cartesian coordinate system. The cube array is composed of M_x, M_y and M_z omnidirectional microphones along the x-axis, y-axis and z-axis, respectively. We denote the positions of the microphones by $(m_\mathrm{x}, m_\mathrm{y}, m_\mathrm{z})$, with $m_\mathrm{x} = 1, 2, \ldots, M_\mathrm{x}$, $m_\mathrm{y} = 1, 2, \ldots, M_\mathrm{y}$ and $m_\mathrm{z} = 1, 2, \ldots, M_\mathrm{z}$. Then, defining the microphone located in $(1, 1, 1)$ as the origin of the coordinate system, the array steering vector of length $M_\mathrm{x} M_\mathrm{y} M_\mathrm{z}$ is expressed by (Van Trees, 2004):

$$\mathbf{d}_{\theta,\phi}(\omega) = \mathbf{c}_\theta(\omega) \otimes \mathbf{b}_{\theta,\phi}(\omega) \otimes \mathbf{a}_{\theta,\phi}(\omega), \tag{1}$$

where

$$\mathbf{a}_{\theta,\phi}(\omega) = \begin{bmatrix} 1 & e^{\jmath \varpi_{\theta,\phi,\mathrm{x}}(\omega)} & \cdots & e^{\jmath (M_\mathrm{x}-1) \varpi_{\theta,\phi,\mathrm{x}}(\omega)} \end{bmatrix}^T \tag{2}$$

is the steering vector associated with the x-axis,

$$\mathbf{b}_{\theta,\phi}(\omega) = \begin{bmatrix} 1 & e^{\jmath \varpi_{\theta,\phi,\mathrm{y}}(\omega)} & \cdots & e^{\jmath (M_\mathrm{y}-1) \varpi_{\theta,\phi,\mathrm{y}}(\omega)} \end{bmatrix}^T \tag{3}$$

is the steering vector associated with the y-axis,

$$\mathbf{c}_{\theta}(\omega) = \begin{bmatrix} 1 & e^{\jmath \varpi_{\theta,\mathrm{z}}(\omega)} & \cdots & e^{\jmath (M_\mathrm{z}-1) \varpi_{\theta,\mathrm{z}}(\omega)} \end{bmatrix}^T \tag{4}$$

is the steering vector associated with the z-axis,

$$\varpi_{\theta,\phi,\mathrm{x}}(\omega) = \frac{\omega \delta_\mathrm{x} \sin\theta \cos\phi}{c},$$
$$\varpi_{\theta,\phi,\mathrm{y}}(\omega) = \frac{\omega \delta_\mathrm{y} \sin\theta \sin\phi}{c},$$
$$\varpi_{\theta,\mathrm{z}}(\omega) = \frac{\omega \delta_\mathrm{z} \cos\theta}{c},$$

δ_x, δ_y, and δ_z are interelement spacings along the x-axis, y-axis and the z-axis, respectively, the superscript ^{T} denotes the transpose operator, $\otimes$ is the KP operator, $\jmath = \sqrt{-1}$ is the imaginary unit, $\omega = 2\pi f$ is the angular frequency, and $f > 0$ is the temporal frequency.

Exploiting the steering vector in (1), the observed signal vector of length $M_\mathrm{x} M_\mathrm{y} M_\mathrm{z}$ of the cube array can be expressed in the frequency domain as (Benesty et al., 2018):

$$\begin{aligned} \mathbf{y}(\omega) &= \begin{bmatrix} \mathbf{y}_1^T(\omega) & \mathbf{y}_2^T(\omega) & \cdots & \mathbf{y}_{M_\mathrm{z}}^T(\omega) \end{bmatrix}^T \\ &= \mathbf{x}(\omega) + \mathbf{v}(\omega) \\ &= \mathbf{d}_{\theta,\phi}(\omega) X(\omega) + \mathbf{v}(\omega), \end{aligned} \tag{5}$$

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

$$\begin{aligned} \mathbf{y}_{m_\mathrm{z}}(\omega) &= \begin{bmatrix} \mathbf{y}_{m_\mathrm{z},1}^T(\omega) & \mathbf{y}_{m_\mathrm{z},2}^T(\omega) & \cdots & \mathbf{y}_{m_\mathrm{z},M_\mathrm{y}}^T(\omega) \end{bmatrix}^T \\ &= \mathbf{x}_{m_\mathrm{z}}(\omega) + \mathbf{v}_{m_\mathrm{z}}(\omega), \end{aligned} \tag{6}$$

is the observed signal vector of length $M_\mathrm{x} \times M_\mathrm{y}$ for a given value of m_z and

$$\begin{aligned} \mathbf{y}_{m_\mathrm{z},m_\mathrm{y}}(\omega) &= \begin{bmatrix} Y_{m_\mathrm{z},m_\mathrm{y},1}(\omega) & \cdots & Y_{m_\mathrm{z},m_\mathrm{y},M_\mathrm{x}}(\omega) \end{bmatrix}^T \\ &= \mathbf{x}_{m_\mathrm{z},m_\mathrm{y}}(\omega) + \mathbf{v}_{m_\mathrm{z},m_\mathrm{y}}(\omega), \end{aligned} \tag{7}$$

is the observed signal vector of length M_x for given values of m_z and m_y. Denoting the desired signal incident angle by (θ_0, ϕ_0) and dropping the dependence on ω, (5) becomes:

$$\mathbf{y} = \left(\mathbf{c}_{\theta_0} \otimes \mathbf{b}_{\theta_0,\phi_0} \otimes \mathbf{a}_{\theta_0,\phi_0} \right) X + \mathbf{v}, \tag{8}$$

where $\mathbf{c}_{\theta_0} \otimes \mathbf{b}_{\theta_0,\phi_0} \otimes \mathbf{a}_{\theta_0,\phi_0} = \mathbf{d}_{\theta_0,\phi_0}$ is the steering matrix at (θ_0, ϕ_0), and the covariance matrix of $\mathbf{y}$ is:

$$\mathbf{\Phi}_\mathbf{y} = E\left(\mathbf{y}\mathbf{y}^H\right) = \phi_X \mathbf{d}_{\theta_0,\phi_0} \mathbf{d}_{\theta_0,\phi_0}^H + \mathbf{\Phi}_\mathbf{v}, \tag{9}$$

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 (9) as:

$$\mathbf{\Phi}_\mathbf{y} = p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0} + p_V \mathbf{\Gamma}_\mathbf{v}, \tag{10}$$

where p_V is the variance of the noise at the reference microphone (i.e., the origin of the Cartesian coordinate system) and $\mathbf{\Gamma}_\mathbf{v} = \mathbf{\Phi}_\mathbf{v}/p_V$ is the pseudo-coherence matrix of the noise. From (10), we deduce that the input signal-to-noise ratio (SNR) is:

$$\mathrm{iSNR} = \frac{\mathrm{tr}\left(p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0}\right)}{\mathrm{tr}\left(p_V \mathbf{\Gamma}_\mathbf{v}\right)} = \frac{p_X}{p_V}, \tag{11}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a square matrix. In the case of the spherically isotropic (diffuse) noise field, (10) becomes:

$$\mathbf{\Phi}_\mathbf{y} = p_X \mathbf{d}_{\theta_0,\phi_0}\mathbf{d}^H_{\theta_0,\phi_0} + p_V \mathbf{\Gamma}_\mathrm{d}, \tag{12}$$

where p_V is the variance of the diffuse noise and $\mathbf{\Gamma}_\mathrm{d}$ is the $M_\mathrm{x}M_\mathrm{y}M_\mathrm{z} \times M_\mathrm{x}M_\mathrm{y}M_\mathrm{z}$ pseudo-coherence matrix of the diffuse noise. We have

$$\mathbf{\Gamma}_\mathrm{d} = \begin{bmatrix} \mathbf{\Gamma}_{\mathrm{d};1} & \cdots & \mathbf{\Gamma}_{\mathrm{d};M_\mathrm{z}-1} & \mathbf{\Gamma}_{\mathrm{d};M_\mathrm{z}} \\ \mathbf{\Gamma}_{\mathrm{d};2} & \cdots & \mathbf{\Gamma}_{\mathrm{d};M_\mathrm{z}-2} & \mathbf{\Gamma}_{\mathrm{d};M_\mathrm{z}-1} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{\Gamma}_{\mathrm{d};M_\mathrm{z}-1} & \cdots & \mathbf{\Gamma}_{\mathrm{d};1} & \mathbf{\Gamma}_{\mathrm{d};2} \\ \mathbf{\Gamma}_{\mathrm{d};M_\mathrm{z}} & \cdots & \mathbf{\Gamma}_{\mathrm{d};2} & \mathbf{\Gamma}_{\mathrm{d};1} \end{bmatrix}, \tag{13}$$

where

$$\mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z}} = \begin{bmatrix} \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},1} & \cdots & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},M_\mathrm{y}-1} & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},M_\mathrm{y}} \\ \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},2} & \cdots & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},M_\mathrm{y}-2} & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},M_\mathrm{y}-1} \\ \vdots & \ddots & \vdots & \vdots \\ \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},M_\mathrm{y}-1} & \cdots & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},1} & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},2} \\ \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},M_\mathrm{y}} & \cdots & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},2} & \mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},1} \end{bmatrix} \tag{14}$$

is a symmetric block Toeplitz matrix of size $M_\mathrm{x}M_\mathrm{y} \times M_\mathrm{x}M_\mathrm{y}$, $m_\mathrm{z} = 1, 2, \ldots, M_\mathrm{z}$, and the elements of the M_y symmetric Toeplitz matrices $\mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},m_\mathrm{y}}$, $m_\mathrm{y} = 1, 2, \ldots, M_\mathrm{y}$ (of size $M_\mathrm{x} \times M_\mathrm{x}$) are given by (Van Trees, 2004):

$$\left(\mathbf{\Gamma}_{\mathrm{d};m_\mathrm{z},m_\mathrm{y}}\right)_{ij} = \mathrm{sinc}\left[\frac{\omega\sqrt{(i-j)^2\delta_\mathrm{x}^2 + \left(m_\mathrm{y}-1\right)^2\delta_\mathrm{y}^2 + \left(m_\mathrm{z}-1\right)^2\delta_\mathrm{z}^2}}{c}\right], \tag{15}$$

with $i, j = 1, 2, \ldots, M_\mathrm{x}$ and $\mathrm{sinc}(x) = \sin x/x$.

3. Kronecker-product beamforming

We would like to design a global cube beamformer $\mathbf{f}$ of length $M_\mathrm{x}M_\mathrm{y}M_\mathrm{z}$ as a KP of three linear sub-beamformers designed with respect to each one of the axes of the cube array. Hence, $\mathbf{f}$ is of the form:

$$\mathbf{f} = \mathbf{u} \otimes \mathbf{w} \otimes \mathbf{h}, \tag{16}$$

where $\mathbf{h}$ is a differential sub-beamformer of length M_x, $\mathbf{w}$ is a constant beamwidth sub-beamformer of length M_y and $\mathbf{u}$ is a constant beamwidth sub-beamformer of length M_z. Such a structure enables a flexible design in which three independent criteria or array attributes may be considered simultaneously. Due to the cube array's 3-D geometry, $\mathbf{f}$ is a typically long beamformer. Then, the beamformer output signal is:

$$Z = \mathbf{f}^H\mathbf{y} = X_\mathrm{fd} + V_\mathrm{rn}, \tag{17}$$

where Z is the estimate of X,

$$X_\mathrm{fd} = \left(\mathbf{u}^H\mathbf{c}_{\theta_0,\phi_0}\right)\left(\mathbf{w}^H\mathbf{b}_{\theta_0,\phi_0}\right)\left(\mathbf{h}^H\mathbf{a}_{\theta_0,\phi_0}\right)X \tag{18}$$

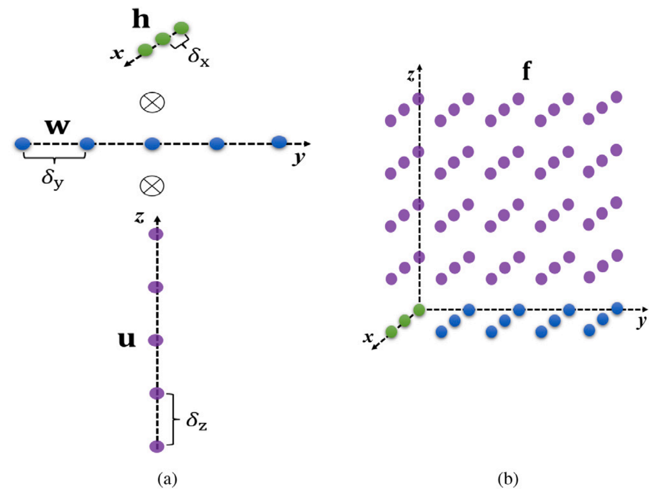

Fig. 1. Illustration of the proposed global cube differential CB beamformer $\mathbf{f}$ as well as the three linear sub-beamformers it is composed of: $\mathbf{h}$, $\mathbf{w}$ and $\mathbf{u}$. (a) The three linear sub-beamformers and (b) the global cube differential CB beamformer. Illustrated array size is $(M_\mathrm{x}, M_\mathrm{y}, M_\mathrm{z}) = (3, 5, 5)$.

is the filtered desired signal, and

$$V_\mathrm{rn} = (\mathbf{u} \otimes \mathbf{w} \otimes \mathbf{h})^H\,\mathbf{v} \tag{19}$$

is the residual noise. As a result, the variance of Z is:

$$\phi_Z = \mathbf{f}^H\mathbf{\Phi}_\mathbf{y}\mathbf{f} = \phi_{X_\mathrm{fd}} + \phi_{V_\mathrm{rn}}, \tag{20}$$

where $\phi_{X_\mathrm{fd}} = p_X\left|\mathbf{u}^H\mathbf{c}_{\theta_0,\phi_0}\right|^2\left|\mathbf{w}^H\mathbf{b}_{\theta_0,\phi_0}\right|^2\left|\mathbf{h}^H\mathbf{a}_{\theta_0,\phi_0}\right|^2$ and $\phi_{V_\mathrm{rn}} = (\mathbf{u} \otimes \mathbf{w} \otimes \mathbf{h})^H\,\mathbf{\Phi}_\mathbf{v}\,(\mathbf{u} \otimes \mathbf{w} \otimes \mathbf{h})$. In addition, it is clear that a distortionless constraint is given by (see Fig. 1):

$$\mathbf{h}^H\mathbf{a}_{\theta_0,\phi_0} = 1,\ \mathbf{w}^H\mathbf{b}_{\theta_0,\phi_0} = 1,\ \mathbf{u}^H\mathbf{c}_{\theta_0,\phi_0} = 1. \tag{21}$$

Next, we relate the most prominent performance measures corresponding to $\mathbf{f}$. The output SNR and the gain in SNR are, respectively,

$$\mathrm{oSNR}(\mathbf{f}) = \frac{p_X}{p_V} \times \frac{\left|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}\right|^2}{\mathbf{f}^H\mathbf{\Gamma}_\mathbf{v}\mathbf{f}}, \tag{22}$$

and

$$\mathcal{G}(\mathbf{f}) = \frac{\mathrm{oSNR}(\mathbf{f})}{\mathrm{iSNR}} = \frac{\left|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}\right|^2}{\mathbf{f}^H\mathbf{\Gamma}_\mathbf{v}\mathbf{f}}, \tag{23}$$

from which we deduce the WNG:

$$\begin{aligned}\mathcal{W}(\mathbf{f}) &= \frac{\left|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}\right|^2}{\mathbf{f}^H\mathbf{f}} \\ &= \frac{\left|\mathbf{u}^H\mathbf{c}_{\theta_0,\phi_0}\right|^2}{\mathbf{c}^H\mathbf{u}} \times \frac{\left|\mathbf{w}^H\mathbf{b}_{\theta_0,\phi_0}\right|^2}{\mathbf{w}^H\mathbf{w}} \times \frac{\left|\mathbf{h}^H\mathbf{a}_{\theta_0,\phi_0}\right|^2}{\mathbf{h}^H\mathbf{h}} \\ &= \mathcal{W}(\mathbf{u}) \times \mathcal{W}(\mathbf{w}) \times \mathcal{W}(\mathbf{h}),\end{aligned} \tag{24}$$

and the DF:

$$\mathcal{D}(\mathbf{f}) = \frac{\left|\mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0}\right|^2}{\mathbf{f}^H\mathbf{\Gamma}_\mathrm{d}\mathbf{f}}. \tag{25}$$

We end by defining the beampattern by:

$$\begin{aligned}\mathcal{B}_{\theta,\phi}(\mathbf{f}) &= \mathbf{f}^H\mathbf{d}_{\theta,\phi} \\ &= \left(\mathbf{u}^H\mathbf{c}_{\theta,\phi}\right)\left(\mathbf{w}^H\mathbf{b}_{\theta,\phi}\right)\left(\mathbf{h}^H\mathbf{a}_{\theta,\phi}\right) \\ &= \mathcal{B}_{\theta,\phi}(\mathbf{u})\,\mathcal{B}_{\theta,\phi}(\mathbf{w})\,\mathcal{B}_{\theta,\phi}(\mathbf{h}),\end{aligned} \tag{26}$$

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

4. Optimal constant-beamwidth beamforming

Assume we are interested in deriving a cube version of the super-directive beamformer (Benesty et al., 2018), which is not necessarily a KP beamformer. That is, we would like to solve

$$\min_{\mathbf{f}} \ \mathbf{f}^H \boldsymbol{\Gamma}_{\mathrm{d}} \mathbf{f} \quad \text{s. t.} \quad \mathbf{f}^H \mathbf{d}_{\theta_0,\phi_0} = 1, \tag{27}$$

whose solution is obtained by

$$\begin{aligned}\mathbf{f}_{\mathrm{SD}} &= \frac{\boldsymbol{\Gamma}_{\mathrm{d}}^{-1} \mathbf{d}_{\theta_0,\phi_0}}{\mathbf{d}_{\theta_0,\phi_0}^H \boldsymbol{\Gamma}_{\mathrm{d}}^{-1} \mathbf{d}_{\theta_0,\phi_0}} \\ &= \frac{\left[\Sigma_{q=1}^{M_{\mathrm{z}}-1} \Sigma_{p=1}^{M_{\mathrm{y}}-1} \mathbf{J}_{M_{\mathrm{z}},q} \otimes \mathbf{J}_{M_{\mathrm{y}},p} \otimes \boldsymbol{\Gamma}_{\mathrm{d};q,p}\right]^{-1} \mathbf{d}_{\theta_0,\phi_0}}{\mathbf{d}_{\theta_0,\phi_0}^H \boldsymbol{\Gamma}_{\mathrm{d}}^{-1} \mathbf{d}_{\theta_0,\phi_0}},\end{aligned} \tag{28}$$

where

$$\left(\mathbf{J}_{M_{\mathrm{y}},p}\right)_{ij} = \begin{cases} 1 & |i-j| = p \\ 0 & |i-j| \neq p \end{cases}, \tag{29}$$

is a binary matrix of size $M_{\mathrm{y}} \times M_{\mathrm{y}}$ with ones on the $-p$th and pth diagonals and zeros elsewhere, and $\mathbf{J}_{M_{\mathrm{z}},q}$ is a binary matrix of size $M_{\mathrm{z}} \times M_{\mathrm{z}}$ defined similarly. In particular, $\mathbf{J}_{M_{\mathrm{y}},0} = \mathbf{I}_{M_{\mathrm{y}}}$ and $\mathbf{J}_{M_{\mathrm{z}},0} = \mathbf{I}_{M_{\mathrm{z}}}$, which are the identity matrices of size $M_{\mathrm{y}} \times M_{\mathrm{y}}$ and $M_{\mathrm{z}} \times M_{\mathrm{z}}$, respectively. Now, assuming

$$\begin{aligned}&\frac{\delta_{\mathrm{y}}}{\delta_{\mathrm{x}}} > M_{\mathrm{x}} - 1, \\ &\frac{\delta_{\mathrm{z}}}{\delta_{\mathrm{x}}} > M_{\mathrm{x}} - 1,\end{aligned} \tag{30}$$

(28) may be approximated by

$$\begin{aligned}\mathbf{f}_{\mathrm{SD}} &\approx \kappa \left(\mathbf{I}_{M_{\mathrm{z}}} \otimes \mathbf{I}_{M_{\mathrm{y}}} \otimes \boldsymbol{\Gamma}_{\mathrm{d};1,1}^{-1}\right) \mathbf{d}_{\theta_0,\phi_0} \\ &= \kappa \left(\mathbf{I}_{M_{\mathrm{z}}} \otimes \mathbf{I}_{M_{\mathrm{y}}} \otimes \boldsymbol{\Gamma}_{\mathrm{d};1,1}^{-1}\right) \left(\mathbf{c}_{\theta_0,\phi_0} \otimes \mathbf{b}_{\theta_0,\phi_0} \otimes \mathbf{a}_{\theta_0,\phi_0}\right) \\ &= \kappa \left(\mathbf{I}_{M_{\mathrm{z}}} \mathbf{c}_{\theta_0,\phi_0}\right) \otimes \left(\mathbf{I}_{M_{\mathrm{y}}} \mathbf{b}_{\theta_0,\phi_0}\right) \otimes \left(\boldsymbol{\Gamma}_{\mathrm{d};1,1}^{-1} \mathbf{a}_{\theta_0,\phi_0}\right) \\ &= \kappa \bar{\mathbf{u}}_{\mathrm{DS}} \otimes \bar{\mathbf{w}}_{\mathrm{DS}} \otimes \bar{\mathbf{h}}_{\mathrm{SD}},\end{aligned} \tag{31}$$

where κ constitutes a normalization factor, $\bar{\mathbf{w}}_{\mathrm{DS}}$ and $\bar{\mathbf{u}}_{\mathrm{DS}}$ are the (unnormalized) linear DS beamformers which operate on the y and z axes, respectively, and $\bar{\mathbf{h}}_{\mathrm{SD}}$ is the (unnormalized) linear SD beamformer which operates on the x-axis. In addition, the condition in (30) implies that the latter should be designed as differential beamformers, that is, with a small interelement spacing δ_{x}, whereas the interelement spacing along the y and z axes, δ_{y} and δ_{z}, should be larger. We note that the approximation in (31) is particularly more accurate in higher frequencies. An illustration of the proposed global cube beamformer is depicted in Fig. 1.

Since the optimal cube SD beamformer can be decomposed into a KP of three linear beamformers, with merely one of which optimized with respect to the array directivity, we may adapt the complementary beamformer to attain another array attribute. For example, to obtain constant-beamwidth beamformers, $\bar{\mathbf{w}}_{\mathrm{DS}}$ and $\bar{\mathbf{u}}_{\mathrm{DS}}$ may be replaced by the modified rectangular window beamformer of Rosen et al. (2017). Furthermore, $\bar{\mathbf{w}}$ and $\bar{\mathbf{u}}$ may be designed as any of the linear window-based constant-beamwidth beamformers suggested in Long et al. (2019). Assuming the desired signal (speaker) is located on the x − y plane in the endfire direction, i.e., $\theta_0 = \pi/2, \phi_0 = 0$, this implies that

$$\begin{aligned}\mathbf{f}_{\mathrm{SD/CB/CB}} &= \kappa \mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}} \otimes \bar{\mathbf{h}}_{\mathrm{SD}} \\ &= \frac{\mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}} \otimes \left(\boldsymbol{\Gamma}_{\mathrm{d};1,1}^{-1} \mathbf{a}_{\pi/2,0}\right)}{\mathbf{a}_{\pi/2,0}^H \boldsymbol{\Gamma}_{\mathrm{d};1,1}^{-1} \mathbf{a}_{\pi/2,0}} \\ &= \mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}} \otimes \mathbf{h}_{\mathrm{SD}},\end{aligned} \tag{32}$$

with $\mathbf{w}_{\mathrm{rect}}$ and $\mathbf{u}_{\mathrm{rect}}$ being the taps of the (normalized) linear modified rectangular window-based constant-beamwidth sub-beamformers of lengths M_{y} and M_{z}, respectively,

$$\mathbf{h}_{\mathrm{SD}} = \frac{\boldsymbol{\Gamma}_{\mathrm{d};1,1}^{-1} \mathbf{a}_{\pi/2,0}}{\mathbf{a}_{\pi/2,0}^H \boldsymbol{\Gamma}_{\mathrm{d};1,1}^{-1} \mathbf{a}_{\pi/2,0}}, \tag{33}$$

and

$$\mathbf{a}_{\pi/2,0} = \begin{bmatrix} 1 & e^{\jmath\omega\delta_{\mathrm{x}}/c} & \cdots & e^{\jmath\omega(M_{\mathrm{x}}-1)\delta_{\mathrm{x}}/c} \end{bmatrix}^T. \tag{34}$$

In many cases, it is desirable to explicitly set either the WNG or DF of the global beamformer. Therefore, we take advantage of the KP beamforming structure, and the approach suggested in Berkun et al. (2015) in the context of ULAs, and modify $\mathbf{h}_{\mathrm{SD}}$ accordingly.

Let us start with the WNG measure and let $\mathcal{W}_0$ be a desirable frequency-dependent WNG value of the global beamformer. Noting that

$$\mathbf{f}_{\mathrm{SD/CB/CB}}^H \boldsymbol{\Gamma}_{\mathrm{d}} \mathbf{f}_{\mathrm{SD/CB/CB}} = \mathbf{h}_{\mathrm{SD}}^H \boldsymbol{\Gamma}_{\mathrm{d,uw}} \mathbf{h}_{\mathrm{SD}}, \tag{35}$$

where

$$\boldsymbol{\Gamma}_{\mathrm{d,uw}} = \left(\mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}} \otimes \mathbf{I}_{M_{\mathrm{x}}}\right)^H \boldsymbol{\Gamma}_{\mathrm{d}} \left(\mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}} \otimes \mathbf{I}_{M_{\mathrm{x}}}\right), \tag{36}$$

we may exploit the approach suggested in Berkun et al. (2015) and define the tunable super-directive beamformer by

$$\mathbf{h}_{\mathrm{SD},\epsilon} = \frac{\left[\boldsymbol{\Gamma}_{\mathrm{d,uw},\epsilon}^{-1} + \alpha \mathbf{I}_{M_{\mathrm{x}}}\right] \mathbf{a}_{\pi/2,0}}{\mathbf{a}_{\pi/2,0}^H \left[\boldsymbol{\Gamma}_{\mathrm{d,uw},\epsilon}^{-1} + \alpha \mathbf{I}_{M_{\mathrm{x}}}\right] \mathbf{a}_{\pi/2,0}}, \tag{37}$$

where

$$\alpha = \frac{\mathbf{a}_{\pi/2,0}^H \boldsymbol{\Gamma}_{\mathrm{d,uw},\epsilon}^{-1} \mathbf{a}_{\pi/2,0}}{M_{\mathrm{x}}} \left[\sqrt{\frac{\bar{\mathcal{W}}_0}{M_{\mathrm{x}} - \bar{\mathcal{W}}_0}} \left|\tan \varphi_\epsilon\right| - 1\right], \tag{38}$$

$$\boldsymbol{\Gamma}_{\mathrm{d,uw},\epsilon} = \boldsymbol{\Gamma}_{\mathrm{d,uw}} + \epsilon \mathbf{I}_{M_{\mathrm{x}}}, \tag{39}$$

and $\epsilon = 10^{-7}$ being a frequency-independent regularization factor. In addition, $\tan \varphi_\epsilon$ may be extracted from

$$\cos \varphi_\epsilon = \frac{\mathbf{a}_{\pi/2,0}^H \boldsymbol{\Gamma}_{\mathrm{d,uw},\epsilon}^{-1} \mathbf{a}_{\pi/2,0}}{\sqrt{M_{\mathrm{x}}} \sqrt{\mathbf{a}_{\pi/2,0}^H \boldsymbol{\Gamma}_{\mathrm{d,uw},\epsilon}^{-2} \mathbf{a}_{\pi/2,0}}}, \tag{40}$$

and $\bar{\mathcal{W}}_0$ is given by

$$\bar{\mathcal{W}}_0 = \mathcal{W}_0 / \mathcal{W}\left(\mathbf{w}_{\mathrm{rect}}\right) / \mathcal{W}\left(\mathbf{u}_{\mathrm{rect}}\right). \tag{41}$$

Then, the WNG of

$$\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon} = \mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}} \otimes \mathbf{h}_{\mathrm{SD},\epsilon} \tag{42}$$

is guaranteed to be $\mathcal{W}_0$. Clearly, invoking (24), we have

$$\begin{aligned}\mathcal{W}_0 &\leq \mathcal{W}\left(\mathbf{u}_{\mathrm{rect}}\right) \times \mathcal{W}\left(\mathbf{w}_{\mathrm{rect}}\right) \times \max\ \bar{\mathcal{W}}_0 \\ &= \mathcal{W}\left(\mathbf{u}_{\mathrm{rect}}\right) \times \mathcal{W}\left(\mathbf{w}_{\mathrm{rect}}\right) \times M_{\mathrm{x}} \\ &= \mathcal{W}_{\mathrm{max,uw}},\end{aligned} \tag{43}$$

where

$$\bar{\mathcal{W}}_0 \leq M_{\mathrm{x}}, \tag{44}$$

with its maximum obtained for $\alpha \longrightarrow \infty$ as $\mathbf{h}_{\mathrm{SD},\epsilon} \longrightarrow \mathbf{h}_{\mathrm{DS}}$, and $\mathbf{h}_{\mathrm{DS}}$ is the DS beamformer.

Similarly, we may wish to set a desirable DF level of the global beamformer by substituting $\boldsymbol{\Gamma}_{\mathrm{d}}$ with $\boldsymbol{\Gamma}_{\mathrm{d,uw}}$ in equations (57)–(59) in Berkun et al. (2015) to obtain $\widetilde{\mathbf{h}}_{\mathrm{SD},\epsilon}$. Then, it is straightforward to show that the DF of the following beamformer

$$\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon} = \mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}} \otimes \widetilde{\mathbf{h}}_{\mathrm{SD},\epsilon}, \tag{45}$$

equals a desirable frequency-dependent value $\mathcal{D}_0$. Clearly, we have

$$\mathcal{D}_0 \leq \left[\widetilde{\mathbf{h}}^H_{\mathrm{SD},\epsilon,\max}\mathbf{\Gamma}_{\mathrm{d},\mathbf{uw},\epsilon}\widetilde{\mathbf{h}}_{\mathrm{SD},\epsilon,\max}\right]^{-1} = \mathcal{D}_{\max,\mathbf{uw},\epsilon}, \tag{46}$$

where

$$\widetilde{\mathbf{h}}_{\mathrm{SD},\epsilon,\max} = \frac{\mathbf{\Gamma}^{-1}_{\mathrm{d},\mathbf{uw},\epsilon}\mathbf{a}_{\pi/2,0}}{\mathbf{a}^H_{\pi/2,0}\mathbf{\Gamma}^{-1}_{\mathrm{d},\mathbf{uw},\epsilon}\mathbf{a}_{\pi/2,0}}. \tag{47}$$

Let us now assume we are interested in deriving 2-D beamformers whose DF is less important than the azimuth and elevation constant-beamwidth property and the WNG. Hence, we set $M_\mathrm{x} = 1$, which implies that the designed beamformers in this part are not differential. We begin by deriving the global optimal rectangular beamformer in terms of the WNG, and as before, we do not restrict the solution to KP beamformers solely. That is, we are interested in solving

$$\min_{\mathbf{f}} \ \mathbf{f}^H\mathbf{f} \quad \text{s. t.} \quad \mathbf{f}^H\mathbf{d}_{\theta_0,\phi_0} = 1. \tag{48}$$

The solution is obtained by

$$\begin{aligned}\mathbf{f}_{\mathrm{DS}} &= \frac{\mathbf{d}_{\theta_0,\phi_0}}{\mathbf{d}^H_{\theta_0,\phi_0}\mathbf{d}_{\theta_0,\phi_0}} \\ &= \frac{\mathbf{c}_{\theta_0,\phi_0}}{\mathbf{c}^H_{\theta_0,\phi_0}\mathbf{c}_{\theta_0,\phi_0}} \otimes \frac{\mathbf{b}_{\theta_0,\phi_0}}{\mathbf{b}^H_{\theta_0,\phi_0}\mathbf{b}_{\theta_0,\phi_0}} \\ &= \mathbf{u}_{\mathrm{DS}} \otimes \mathbf{w}_{\mathrm{DS}},\end{aligned} \tag{49}$$

which is axis-separable without any further approximations. As with the SD beamformer, we may now replace any of the linear DS beamformers with a linear constant-beamwidth beamformer. For example, we may employ the modified rectangular window-based constant-beamwidth beamformer along the y- and z-axes. This implies that Eqs. (2) and (3) are modified accordingly. We note that the speaker is still assumed to be located on the (positive) x-axis, that is, $(\theta_0, \phi_0) = (\pi/2, 0)$. We therefore have

$$\mathbf{f}_{\mathrm{CB/CB}} = \mathbf{u}_{\mathrm{rect}} \otimes \mathbf{w}_{\mathrm{rect}}, \tag{50}$$

which is regarded as the rectangular azimuth and elevation constant-beamwidth beamformer. We note that $\mathbf{f}_{\mathrm{CB/CB}}$ may be seen as a special case of $\mathbf{f}_{\mathrm{SD/CB/CB}}$ with $M_\mathrm{x} = 1$, and in contrast to the former case, the interelement spacing δ_y and δ_z may take any value. In particular, neither should be very small as none of the linear beamformers involved is differential.

It is worthwhile addressing the computational complexity of the proposed beamformers, which is typically determined by the number of multiplications required to derive and apply the beamformers. The derivation of the linear CB sub-beamformers is in the order of $\mathcal{O}(1)$ as it merely requires a small finite number of multiplications independent of the number of microphones. This implies that the derivation of $\mathbf{w}_{\mathrm{rect}}$ and $\mathbf{u}_{\mathrm{rect}}$ is of order $\mathcal{O}(1)$. Still, the KP operation applied to generate $\mathbf{f}_{\mathrm{CB/CB}}$ requires $\mathcal{O}(M_\mathrm{y}M_\mathrm{z})$ multiplications (although in practice this may be reduced as most values are constant). On the contrary, the derivation of either $\widetilde{\mathbf{h}}_{\mathrm{SD},\epsilon}$ or $\mathbf{h}_{\mathrm{SD},\epsilon}$ requires $\mathcal{O}(M_\mathrm{x}^3M_\mathrm{y}^2M_\mathrm{z}^2)$ to derive the matrix $\mathbf{\Gamma}_{\mathrm{d},\mathbf{uw}}$ and another $\mathcal{O}(M_\mathrm{x}^3)$ multiplications to compute its inverse. The application of these three beamformers to the noisy observations vector is simply in the order of the length of the beamformer, that is, $\mathcal{O}(M_\mathrm{x}M_\mathrm{y}M_\mathrm{z})$ multiplications (with $M_\mathrm{x} = 1$ for $\mathbf{f}_{\mathrm{CB/CB}}$). For comparison, the derivation of the cube SD beamformer $\mathbf{f}_{\mathrm{SD}}$ is in the order of $\mathcal{O}(M_\mathrm{x}^3M_\mathrm{y}^3M_\mathrm{z}^3)$ due to the computation of $\mathbf{\Gamma}_\mathrm{d}^{-1}$ (Householder, 2013), and the derivation of the cube DS beamformer $\mathbf{f}_{\mathrm{DS}}$ is of order $\mathcal{O}(M_\mathrm{x})$. The application to the noisy observations vector (roughly, without any implementation optimization) is similar to the proposed beamformers, which lower bounds the computational complexity by $\mathcal{O}(M_\mathrm{x}M_\mathrm{y}M_\mathrm{z})$. Table 1 summarizes the computational complexity of the discussed beamformers.

Table 1
Computational complexity comparison of the proposed beamformers and the cube SD and DS beamformers.

$\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$	$\mathcal{O}(M_\mathrm{x}^3M_\mathrm{y}^2M_\mathrm{z}^2)$
$\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$	$\mathcal{O}(M_\mathrm{x}^3M_\mathrm{y}^2M_\mathrm{z}^2)$
$\mathbf{f}_{\mathrm{CB/CB}}$	$\mathcal{O}(M_\mathrm{y}M_\mathrm{z})$
$\mathbf{f}_{\mathrm{SD}}$	$\mathcal{O}(M_\mathrm{x}^3M_\mathrm{y}^3M_\mathrm{z}^3)$
$\mathbf{f}_{\mathrm{DS}}$	$\mathcal{O}(M_\mathrm{x}M_\mathrm{y}M_\mathrm{z})$

We end this part by comparing the proposed approach to the approaches suggested in Wang et al. (2021), and Itzhak and Cohen (2022), respectively. In essence, all three approaches take advantage of the KP beamforming framework to design flexible global beamformers composed of linear sub-beamformers of different kinds. However, while the approach in Wang et al. (2021) a priori considers fully-separable global beamformers and does not directly address the interelement spacing along the three axes (and is arbitrarily taken as identical along all axes), the proposed approach and the approach suggested in Itzhak and Cohen (2022) are shown to resemble the traditional SD beamformer better (implying better array directivity) when certain design conditions are satisfied (30). Hence, they are designed with a differential sub-beamformer along the x-axis whose interelement spacing is small, and wider-spaced sub-beamformers along the y and z axes. In addition, the work in Wang et al. (2021) does not consider some of the array attributes this work considers, e.g., the CB property and direct control of either the WNG or DF. Relating the work presented in Itzhak and Cohen (2022), it merely focuses on the CB property with respect to azimuth angle and takes advantage of URAs. In contrast, the proposed approach strives to attain a constant main-lobe beamwidth with respect to both the azimuth and elevation angles by using cube arrays. Thus, the latter may be regarded as a generalization of the former. Additionally, we propose the $\mathbf{f}_{\mathrm{CB/CB}}$ beamformer, for which no comparable beamformer is suggested in Itzhak and Cohen (2022). Finally, the analysis performed in this work is much deeper, considering the computational complexity of the proposed beamformers as well as desired speech signals impinging on the array from varying incident angles in space and not necessarily from the x − y plane. This will be thoroughly evaluated and discussed in the next section from both the beampattern and main-lobe beamwidth perspectives, as well as the quality and intelligibility of enhanced speech signals.

5. Simulations

5.1. Design examples

We begin by providing some design examples of $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ and $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$. Considering $M = 105$ as a fixed number of microphones in the array, we are interested in investigating the impact of the values of M_x, M_y and M_z on the beamformer's properties. According to the previous part, we first set the values of M_y and M_z; then, we use Rosen et al. (2017) to obtain $\mathbf{w}_{\mathrm{rect}}$ and $\mathbf{u}_{\mathrm{rect}}$, respectively. Next, we design the differential sub-beamformer along the x-axis using either $\mathbf{h}_{\mathrm{SD},\epsilon}$ or $\widetilde{\mathbf{h}}_{\mathrm{SD},\epsilon}$, which are of length $M_\mathrm{x} = M/M_\mathrm{y}/M_\mathrm{z}$, to set either the DF or the WNG of the cube array. Clearly, with $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$, as long as (43) is satisfied we may set $\mathcal{W}_0$ arbitrarily; and with $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$, as long as (46) is satisfied we may set $\mathcal{D}_0$ arbitrarily. Figs. 2 and 3 demonstrate the WNG and DF with $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ and $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$, respectively. In each, we depict the three possible ULAs which serve as a reference, that is, a tunable super-directive beamformer along the x-axis and two CB beamformers along the y and z axes, respectively. In addition, we plot three distinct examples of the proposed cube beamformers. We note that while in Itzhak and Cohen (2022) it was suggested to set $\mathcal{W}_0$ and $\mathcal{D}_0$ as either frequency-invariant values or as reduced values of their respective maximum, in this work we strictly set $\mathcal{W}_0 = \mathcal{W}_{\max,\mathbf{uw}}$ with $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$, which is therefore denoted as $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon,\max}$, whereas

$\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ is designed with $\mathcal{D}_0 = \mathcal{D}_{\max,\mathbf{uw},\epsilon}$ and is therefore denoted as $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon,\max}$. In addition, we set $\delta_{\mathrm{x}} = 5$ mm, $\delta_{\mathrm{y}} = \delta_{\mathrm{z}} = 4$ cm, and the main-lobe beamwidths with respect to both angles are set to $\Delta_\phi = \Delta_\theta = 40°$. Addressing $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$, we observe that, indeed, the higher the value of M_{x} the better the WNG, in particular in high frequencies. The reason for that is double: (a) since $\mathcal{W}_0 = \mathcal{W}_{\max,\mathbf{uw}}$ the sub-beamformer along the x-axis is the WNG-optimal DS beamformer and (b) unlike the CB sub-beamformers the length of the DS sub-beamformer does not reduce as the frequency increases. The latter also explains the DF performance in high frequencies: the higher M_{x}, the larger the array size and its directivity. Addressing $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$, it is clear that the higher M_{x} the better the DF, which is a consequence of setting $\mathcal{D}_0 = \mathcal{D}_{\max,\mathbf{uw},\epsilon}$. On the contrary, and unlike with the other beamformer, the higher M_{x}, the worse the WNG, which is a consequence of the very low WNG nature of the SD beamformer.

Next, in Fig. 4, we analyze the 2-D azimuth and elevation beampatterns with $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ as a function of the frequency. We note that in the 2-D elevation beampatterns, the θ-angle is measured with respect to the $\mathrm{x}-\mathrm{y}$ plane. It is evident that when M is fixed, increasing M_{y} and M_{z} lowers the CB threshold frequency with respect to the azimuth and elevation angles, respectively. In contrast, increasing either of them with the other remains fixed, implies that M_{x} drops, and accordingly, so does the directivity of the global beamformer, which is expressed by significant side lobes. It is worthwhile noting that a similar behavior and tradeoff exist with the $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ beamformer as well. Consequently, addressing both beamformers, we infer that M_{y} and M_{z} determine the CB threshold frequencies, whereas M_{x} sets the WNG and DF measures of the global beamformer: with $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$, increasing M_{x} maximizes the DF but deteriorates the WNG; with $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ increasing M_{x} maximizes the WNG and the DF improves as well. This is particularly emphasized in higher frequencies.

We move on to investigating the $\mathbf{f}_{\mathrm{CB/CB}}$ beamformer, with the WNG and DF of six of its design examples depicted in Fig. 5. We set $M = M_{\mathrm{y}}M_{\mathrm{z}} = 45$ and keep all other settings identical to the previous part. We immediately note that the ULAs, that is, when either M_{y} or M_{z} equal 1, exhibit inferior performance in terms of both measures. In fact, we observe that the closer the values of M_{y} and M_{z} the better the performance. This is a consequence of the larger effective array size of the global rectangular beamformer: as the frequency increases, the effective length (the number of non-zero taps) of each sub-beamformer drops. For any frequency above the CB threshold frequency, this effective length does not depend on the physical number of microphones in the sub-beamformer (that is, M_{y} or M_{z}) but rather it depends on the frequency, the interelement spacing, and the desired main-lobe beamwidth. This means that the more balanced the axes of the rectangular beamformer, the fewer physical microphones are zeroed and the larger the effective array size (for frequencies above the CB threshold frequency). This also explains why both the WNG and DF measures remain identical upon interchanging the values of M_{y} and M_{z}, that is, by effectively rotating the beamformer. Finally, we point out a clear performance drop in edge frequencies in which the effective length of the sub-beamformers drops. This behavior was previously reported in Rosen et al. (2017) and Itzhak and Cohen (2022).

We end this part by addressing the azimuth and elevation beamwidths with the same six variations of the $\mathbf{f}_{\mathrm{CB/CB}}$ beamformer. The beamwidths are depicted in Fig. 6; to compute them we define the zeros of the beampatterns as at least 30 dB weaker than the maximal gain. It is clear that, as with the two cube beamformers, the larger M_{y}, the lower the CB threshold frequency with respect to the azimuth angle, and the larger M_{z}, the lower the CB threshold frequency with respect to the elevation angle. In particular, when either M_{y} or M_{z} equal 1 constant beamwidth is obtained with respect to merely a single angle. We infer that when the constant beamwidth property is of equal importance with respect to both angles, M_{y} and M_{z} should be roughly equal, which is also the optimal choice in terms of the WNG and DF measures. On the contrary, When the CB property is of a higher interest with respect to a single angle, the corresponding sub-beamformer size should be larger. However, this increases the CB threshold frequency with respect to the complementary angle and deteriorates both the WNG and DF.

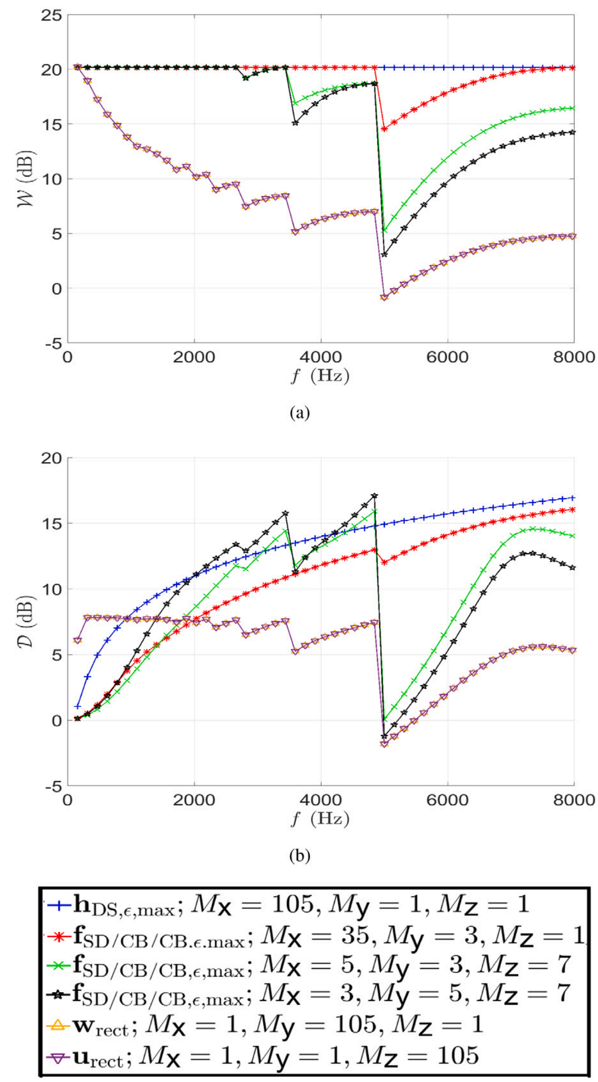

Fig. 2. WNG and DF with $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ for different values of $M_{\mathrm{x}}M_{\mathrm{y}}M_{\mathrm{z}} = 105$. (a) WNG and (b) DF.

5.2. Speech signals simulations in reverberant environments

Next, we analyze the performance of the proposed beamformers with noisy speech signals in reverberant environments and examine their tolerance to deviations in the speech signals' DOA. The reverberant simulations are performed as follows. We use a room impulse response (RIR) generator (Habets, 2008) to simulate the reverberant noise-free signal received in a beamformer consisting of $M_{\mathrm{x}} \times M_{\mathrm{y}} \times M_{\mathrm{z}} = 75$ microphones in four different settings: $(M_{\mathrm{x}}, M_{\mathrm{y}}, M_{\mathrm{z}}) = (3,5,5)$, $(5,3,5)$, $(5,1,15)$ and $(1,5,15)$. For each of the settings, we design each of the three proposed beamformers, that is, $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$, $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ and $\mathbf{f}_{\mathrm{CB/CB}}$, as well as the referred cube SD and DS beamformers, that is, $\mathbf{f}_{\mathrm{SD}}$ and, $\mathbf{f}_{\mathrm{DS}}$, respectively. Additionally, we design the $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG},1}$ and $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG},2}$ beamformers proposed in Wang et al. (2021) with an interelement spacing of $\delta = \delta_{\mathrm{x}} = \delta_{\mathrm{y}} = \delta_{\mathrm{z}} = 1.5$ cm which was found optimal for our scenarios through simulations. While this is a relatively large number of array microphones, it cannot be reduced greatly due to the cube nature of the array. This implies that the proposed approach

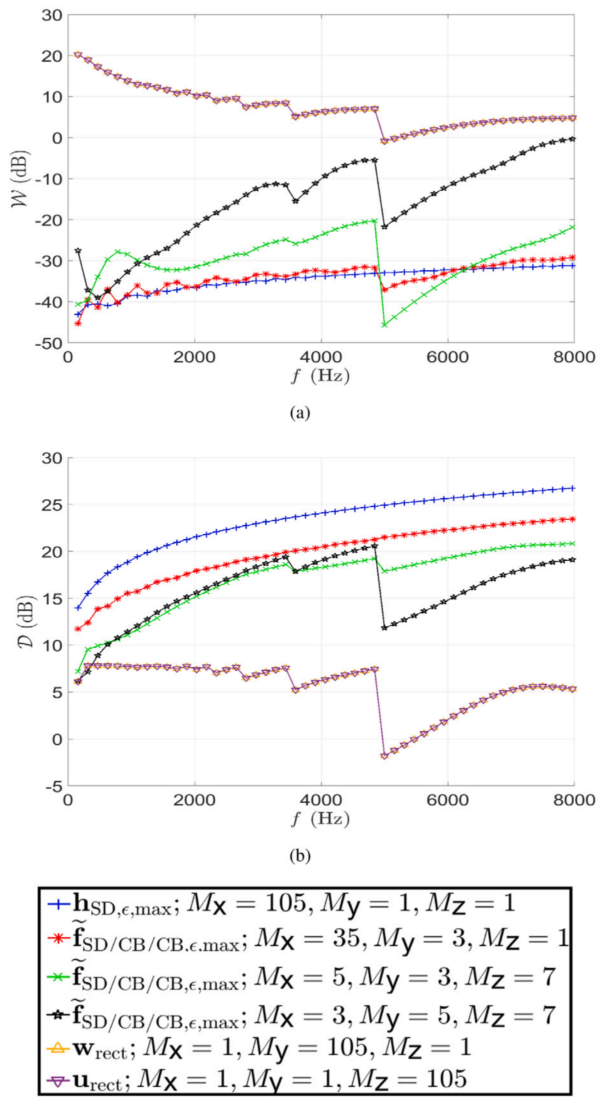

Fig. 3. WNG and DF with $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ for different values of $M_x M_y M_z = 105$. (a) WNG and (b) DF.

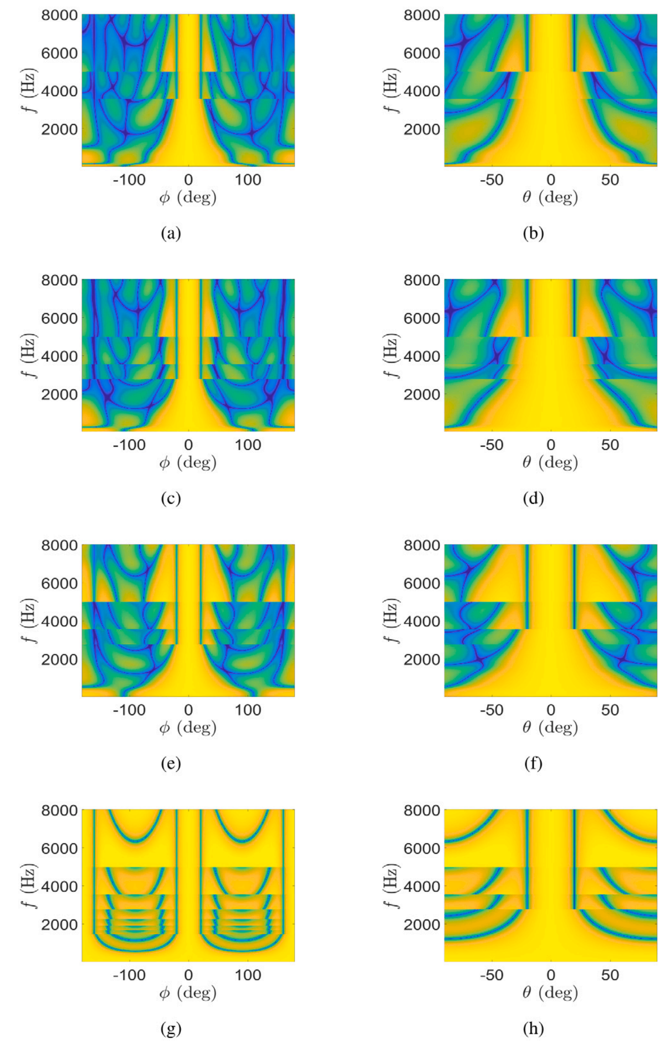

Fig. 4. Azimuth and elevation beampattern analysis with $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ for different values of $M_x M_y M_z = 105$ as a function of the frequency. The θ-angle in the 2-D elevation beampatterns is measured with respect to the x − y plane. (a) Azimuth beampatterns with $M_x = 7, M_y = 5, M_z = 3$, (b) elevation beampatterns with $M_x = 7, M_y = 5, M_z = 3$, (c) azimuth beampatterns with $M_x = 5, M_y = 7, M_z = 3$, (d) elevation beampatterns with $M_x = 5, M_y = 7, M_z = 3$, (e) azimuth beampatterns with $M_x = 3, M_y = 7, M_z = 5$, (f) elevation beampatterns with $M_x = 3, M_y = 7, M_z = 5$, (g) azimuth beampatterns with $M_x = 1, M_y = 15, M_z = 7$, and (h) elevation beampatterns with $M_x = 1, M_y = 15, M_z = 7$.

is less likely to fit physically small or low-cost appliances but could be used in studios, for conference room calls, or potentially any non-portable application. In each of the four settings, the array is located on the $z = 2$ m plane and centered around the $(x, y) = (4\text{ m}, 4\text{ m})$ coordinate, with $\delta_x = 5$ mm and $\delta_y = \delta_z = 4$ cm. The RIR generator is based on the image method of Allen and Berkley (1979). We simulate three distinct scenarios in a 6 × 6 × 5 m room, which differ by the location of the desired speech signal source in the room, which corresponds to three distinct DOAs: (θ_0, ϕ_0): $(\pi/2, 0)$ (no deviation), 5^0 deviation with respect to the azimuth and elevation angles, and a 10^0 deviation with respect to the azimuth and elevation angles. In all scenarios, we set $T_{60} = 250$ msec, where T_{60} is defined by Sabin-Franklin's formula (Pierce, 2019). 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 iSNR = −5 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 (DARPA, 1993). It is sampled at a sampling rate of $f_s = 1/T_s = 16$ kHz within the signal duration T.

The noisy observations signal is transformed into the STFT domain using 75% overlapping time frames and a Hamming analysis window of length 256 (16 msec). The discrete Fourier-transform length is set to 256 as well. Next, rectangular differential beamformers with different configurations are independently applied to the noisy signal to yield clean signal estimates in the STFT domain, followed by an inverse STFT to obtain time-domain enhanced signals.

We analyze and compare the average PESQ (Rix et al., 2001) and STOI (Taal et al., 2011) scores of the time-domain enhanced signals in each of the three aforementioned scenarios and with each of the seven aforementioned beamformers (the three proposed beamformers, the SD beamformer and the DS beamformer, and $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG},1}$ and $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG},2}$ of Wang et al. (2021)). Each beamformer is designed and applied to the noisy observations in each of the four array settings described above. The results are shown in Tables 2–4, respectively, including the PESQ and STOI scores of the noisy signals as received by the reference (first) microphone. We note that each of the scenarios is characterized by a different reverberation pattern (as the location of the speech signal source varies) and hence the scenarios cannot be directly compared but rather individually analyzed. In addition, inapplicable array settings are marked by gray table cells: with $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$

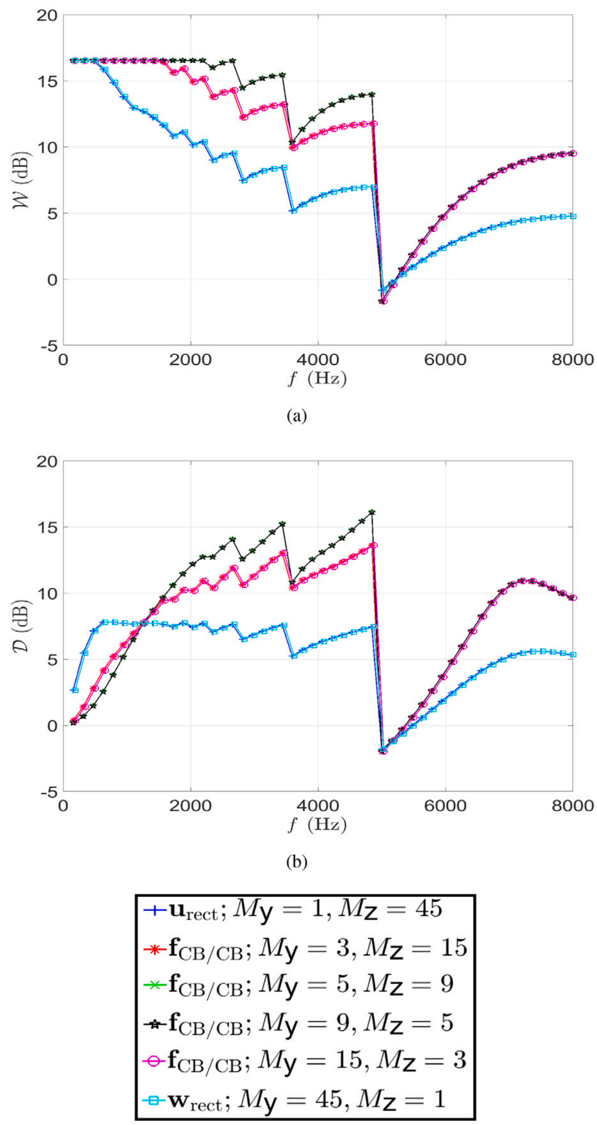

Fig. 5. WNG and DF with $\mathbf{f}_{\mathrm{CB/CB}}$ for different values of $M_y M_z = 45$. (a) WNG and (b) DF.

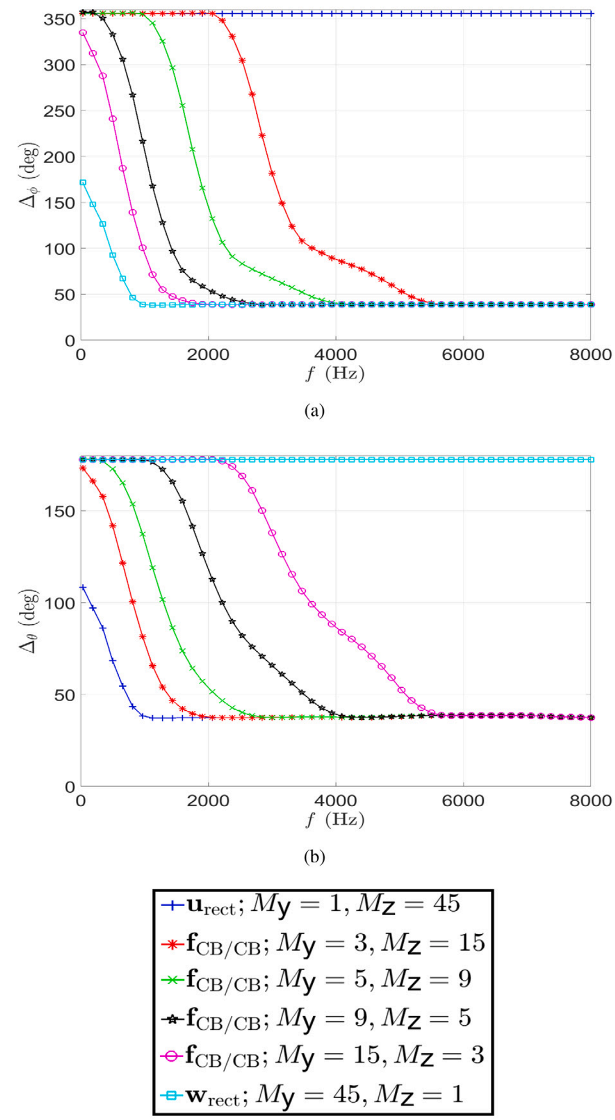

Fig. 6. Azimuth and elevation beamwidths with $\mathbf{f}_{\mathrm{CB/CB}}$ for different values of $M_y M_z = 45$. (a) Azimuth-angle beamwidths and (b) elevation-angle beamwidths.

and $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ we always have $M_x > 1$, whereas with $\mathbf{f}_{\mathrm{CB/CB}}$ we always have $M_x = 1$. To begin with, we observe that without any deviation of the desired signal's DOA (first scenario; Table 2) the $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ beamformer outperforms all other beamformers in terms of both scores. This may be explained by its ability to attenuate both the white noise as well as the diffuse noise and reverberations. This is resulted in by its construction- each of the linear sub-beamformers it is composed of is either optimal (SD; diffuse noise and reverberations) or nearly optimal (CB; white noise) with respect to one of the simulated noise fields. We note that in some sense it is similar by design to $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,1}}$, however, the latter considers an identical interelement spacing along all axes and the linear sub-beamformers it is composed of are independently designed, as opposed to $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ with which $\boldsymbol{\Gamma}_{\mathrm{d,uw}}$ is considered and its interelement spacings along the different axes are not all equal and chosen to comply with (30). In addition, as both M_y and M_z are larger than 1, the $\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$ beamformer exhibits a constant beamwidth with respect to both the azimuth and elevation angles. This may also explain its STOI score performance, which is superior in all three scenarios. Focusing on $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ and $\mathbf{f}_{\mathrm{CB/CB}}$, we observe that both exhibit very similar performance to the DS beamformer $\mathbf{f}_{\mathrm{DS}}$ in terms of both scores. Clearly, this is expected as for every frequency below the CB threshold frequency the linear rectangular window-based sub-beamformers $\mathbf{u}_{\mathrm{rect}}$ and $\mathbf{w}_{\mathrm{rect}}$ converge to the linear DS beamformer whose all taps equal, and for every frequency above the CB threshold frequency merely two taps out of the entire beamformer differ from the rectangular window all-equal taps pattern. Therefore, in future work, we suggest replacing the linear modified rectangular CB beamformers of $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ and $\mathbf{f}_{\mathrm{CB/CB}}$ by an alternative linear CB sub-beamformer (for example, a Kaiser- or a Chebyshev-window based beamformer) whose directivity is potentially preferable and its length does not decrease as the frequency increases. Nevertheless, the approach suggested in Section 4 clearly remains valid regardless of the linear CB sub-beamformer at hand. Addressing the SD beamformer $\mathbf{f}_{\mathrm{SD}}$, we note that it performs well when no DOA deviation exists as it is able to attenuate the diffuse noise and reverberations to the largest extent, however, its STOI scores performance degrades in the two complementary scenarios in which DOA deviations exist (Tables 3 and 4). In some cases, the intelligibility of the enhanced speech signals is even worse than the noisy speech signals. This is a consequence of the frequency-dependent beamwidth of the SD beamformer which considerably distorts the desired signal when its DOA deviates from its expected direction- in particular in high frequencies.

Table 2
Average PESQ and STOI scores of the enhanced signals with the three proposed beamformers, the traditional SD and DS beamformers, and the $\bar{\mathbf{h}}_{\mathrm{aMDF}}$ and $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,1}}$ beamformers proposed in Wang et al. (2021). The simulations are carried out in four distinct array settings (M_x, M_y, M_z), and with no deviation of the desired speech signal's DOA. The PESQ score of the noisy signal is 1.84 and its corresponding STOI score is 0.78.

(M_x, M_y, M_z)	PESQ				STOI			
	(3, 5, 5)	(5, 3, 5)	(5, 1, 15)	(1, 5, 15)	(3, 5, 5)	(5, 3, 5)	(5, 1, 15)	(1, 5, 15)
$\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$	**3.51**	**3.09**	**2.89**		**0.96**	0.86	**0.93**	
$\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$	2.29	2.19	2.10		0.83	0.81	0.79	
$\mathbf{f}_{\mathrm{CB/CB}}$				2.36				0.84
$\mathbf{f}_{\mathrm{SD}}$	2.63	2.21	2.80	**2.56**	0.92	0.86	0.92	**0.90**
$\mathbf{f}_{\mathrm{DS}}$	2.30	2.19	2.17	2.42	0.84	0.82	0.81	0.85
$\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,1}}$ (Wang et al., 2021)	2.54	2.78	2.74	2.08	0.88	**0.92**	0.88	0.80
$\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,2}}$ (Wang et al., 2021)	2.34	2.18	2.00	2.21	0.83	0.77	0.73	0.84

Table 3
Average PESQ and STOI scores of the enhanced signals with the three proposed beamformers, the traditional SD and DS beamformers, and the $\bar{\mathbf{h}}_{\mathrm{aMDF}}$ and $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,1}}$ beamformers proposed in Wang et al. (2021). The simulations are carried out in four distinct array settings (M_x, M_y, M_z), and with a 5° deviation of the desired speech signal's DOA in both the azimuth and elevation angles, respectively. The PESQ score of the noisy signal is 1.74 and its corresponding STOI score is 0.75.

(M_x, M_y, M_z)	PESQ				STOI			
	(3, 5, 5)	(5, 3, 5)	(5, 1, 15)	(1, 5, 15)	(3, 5, 5)	(5, 3, 5)	(5, 1, 15)	(1, 5, 15)
$\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$	**2.34**	2.32	**2.35**		0.72	0.79	**0.87**	
$\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$	2.22	2.09	1.95		0.82	0.80	0.77	
$\mathbf{f}_{\mathrm{CB/CB}}$				2.24				0.82
$\mathbf{f}_{\mathrm{SD}}$	2.10	1.93	2.11	**2.46**	0.70	0.68	0.82	0.71
$\mathbf{f}_{\mathrm{DS}}$	2.21	2.09	2.01	2.30	**0.83**	0.80	0.79	**0.83**
$\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,1}}$ (Wang et al., 2021)	1.93	**2.58**	2.10	2.03	0.81	**0.81**	0.79	0.79
$\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,2}}$ (Wang et al., 2021)	1.89	1.87	1.90	2.13	0.72	0.71	0.71	0.81

Table 4
Average PESQ and STOI scores of the enhanced signals with the three proposed beamformers, the traditional SD and DS beamformers, and the $\bar{\mathbf{h}}_{\mathrm{aMDF}}$ and $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,1}}$ beamformers proposed in Wang et al. (2021). The simulations are carried out in four distinct array settings (M_x, M_y, M_z), and with a 10° deviation of the desired speech signal's DOA in both the azimuth and elevation angles, respectively. The PESQ score of the noisy signal is 1.71 and its corresponding STOI score is 0.70.

(M_x, M_y, M_z)	PESQ				STOI			
	(3, 5, 5)	(5, 3, 5)	(5, 1, 15)	(1, 5, 15)	(3, 5, 5)	(5, 3, 5)	(5, 1, 15)	(1, 5, 15)
$\widetilde{\mathbf{f}}_{\mathrm{SD/CB/CB},\epsilon}$	**2.42**	2.31	**2.29**		0.79	**0.85**	**0.87**	
$\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$	2.22	2.14	2.07		**0.82**	0.79	0.76	
$\mathbf{f}_{\mathrm{CB/CB}}$				2.31				0.81
$\mathbf{f}_{\mathrm{SD}}$	2.10	1.77	2.03	**2.37**	0.72	0.64	0.70	0.81
$\mathbf{f}_{\mathrm{DS}}$	2.22	2.15	2.14	2.35	**0.82**	0.79	0.75	0.80
$\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,1}}$ (Wang et al., 2021)	2.22	**2.33**	1.81	2.09	0.75	0.80	0.72	0.80
$\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,2}}$ (Wang et al., 2021)	1.89	1.94	1.99	2.28	0.73	0.72	0.72	**0.84**

Finally, relating the $\bar{\mathbf{h}}_{\mathrm{aMDF/MWNG,2}}$ beamformer, we note it is generally outperformed by all other beamformers and in all array settings, except when $(M_x, M_y, M_z) = (1, 5, 15)$ in which its STOI scores are comparable (or even superior in the 10^0 deviation scenario).

6. Conclusions

We have presented an azimuth and elevation CB differential beamforming approach for 3-D cube arrays. Assuming the ratios between the interelement spacing along the y and x axes, and the interelement spacing along the z and x axes, are larger than the number of microphones along the x-axis, the cube SD beamformer is shown to be approximated by a KP of a linear DS sub-beamformer along the z-axis, a linear DS sub-beamformer along the y-axis and a linear SD sub-beamformer along the x-axis. Then, we replace the DS sub-beamformers with CB sub-beamformers and propose two design methods to derive global cube beamformers whose either WNG or DF may be set by design. We show that by tuning the values of M_y, M_z and M_x we are able to control the tradeoff between the CB threshold frequency with respect to the azimuth and elevation angles, respectively, and the WNG or DF performance. In addition, we focus on the special case of $M_x = 1$ and show that a rectangular azimuth and elevation CB beamformer may be obtained by applying a KP decomposition to the rectangular DS beamformer and replacing the linear DS sub-beamformers with linear CB sub-beamformers. The azimuth and elevation CB threshold frequencies with this rectangular CB beamformer are controlled by the values of M_y and M_z, respectively, with a lower threshold frequency obtained for a larger corresponding sub-beamformer size. The WNG and DF measures are shown to be maximized when M_y and M_z are equal. Finally, we analyze the performance of the proposed beamformers through speech signals simulations in various reverberant scenarios and array settings, and with deviations of the DOA. We show that the proposed beamformers outperform previously-presented beamformers and the traditional SD and DS beamformers, particularly in terms of the intelligibility of their corresponding time-domain enhanced signals. In future work, we may replace the $\mathbf{w}_{\mathrm{rect}}$ sub-beamformer with another linear CB beamformer to potentially avoid the WNG and DF performance drops in edge frequencies and better differentiate between the proposed $\mathbf{f}_{\mathrm{SD/CB/CB},\epsilon}$ and $\mathbf{f}_{\mathrm{CB/CB}}$ beamformers and the traditional DS beamformer.

CRediT authorship contribution statement

Gal Itzhak: Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Israel Cohen:** Conceptualization, Methodology, Formal analysis, Investigation, Writing – original draft, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article

References

Abramovich, Y.I., Frazer, G.J., Johnson, B.A., 2010. Iterative adaptive Kronecker MIMO radar beamformer: Description and convergence analysis. IEEE Trans. Signal Process. 58 (7), 3681–3691.

Allen, J.B., Berkley, D.A., 1979. Image method for efficiently simulating small-room acoustics. J. Acoust. Soc. Am. 65 (4), 943–950.

Benesty, J., Chen, J., Cohen, I., 2015. Design of Circular Differential Microphone Arrays. Springer-Verlag, Switzerland.

Benesty, J., Cohen, I., Chen, J., 2018. Fundamentals of Signal Enhancement and Array Signal Processing. Wiley-IEEE Press, Singapore.

Benesty, J., Cohen, I., Chen, J., 2019. Array Processing - Kronecker Product Beamforming. Springer-Verlag, Switzerland.

Berkun, R., Cohen, I., Benesty, J., 2015. Combined beamformers for robust broadband regularized superdirective beamforming. IEEE/ACM Trans. Audio Speech Lang. Process. 23 (5), 877–886.

Buchris, Y., Cohen, I., Benesty, J., 2018. Frequency-domain design of asymmetric circular differential microphone arrays. IEEE/ACM Trans. Audio Speech Lang. Process. 26 (4), 760–773.

Buchris, Y., Cohen, I., Benesty, J., 2019. On the design of time-domain differential microphone arrays. Appl. Acoust. 148, 212–222.

Cohen, I., Benesty, J., Chen, J., 2019. Differential Kronecker product beamforming. IEEE/ACM Trans. Audio Speech Lang. Process. 27 (5), 892–902.

1993. DARPA TIMIT acoustic phonetic continuous speech corpus.

Habets, E.A.P., 2008. Room impulse response (RIR) generator.

Heidenreich, P., Zoubir, A.M., Rubsamen, M., 2012. Joint 2-D DOA estimation and phase calibration for uniform rectangular arrays. IEEE Trans. Signal Process. 60 (9), 4683–4693.

Householder, A.S., 2013. The Theory of Matrices in Numerical Analysis. Dover Publications, New York.

Huang, G., Chen, J., Benesty, J., 2018. On the design of differential beamformers with arbitrary planar microphone array geometry. J. Acoust. Soc. Am. 144 (1), EL66–EL70.

Itzhak, G., Benesty, J., Cohen, I., 2019. Nonlinear Kronecker product filtering for multichannel noise reduction. Speech Commun. 114, 49–59.

Itzhak, G., Benesty, J., Cohen, I., 2021. On the design of differential Kronecker product beamformers. IEEE/ACM Trans. Audio Speech Lang. Process. 29, 1397–1410.

Itzhak, G., Benesty, J., Cohen, I., 2021. Quadratic beamforming for magnitude estimation. In: Proc. 29th European Signal Processing Conference. EUSIPCO, pp. 251–255.

Itzhak, G., Benesty, J., Cohen, I., 2022. Multistage approach for steerable differential beamforming with rectangular arrays. Speech Commun. 142, 61–76.

Itzhak, G., Cohen, I., 2022. Differential and constant-beamwidth beamforming with uniform rectangular arrays. In: Proc. 17th International Workshop on Acoustic Signal Enhancement, IWAENC-2022.

Itzhak, G., Cohen, I., Benesty, J., 2021. Robust differential beamforming with rectangular arrays. In: Proc. 29th European Signal Processing Conference, EUSIPCO-2021.

Jin, J., Huang, G., Wang, X., Chen, J., Benesty, J., Cohen, I., 2021. Steering study of linear differential microphone arrays. IEEE/ACM Trans. Audio Speech Lang. Process. 29, 158–170.

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

Khaykin, D., Rafaely, B., 2009. Coherent signals direction-of-arrival estimation using a spherical microphone array: Frequency smoothing approach. In: 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. pp. 221–224.

Konforti, Y., Cohen, I., Berdugo, B., 2022. Array geometry optimization for region-of-interest broadband beamforming. In: Proc. 17th International Workshop on Acoustic Signal Enhancement, IWAENC-2022.

Koretz, A., Rafaely, B., 2009. Dolph–Chebyshev beampattern design for spherical arrays. IEEE Trans. Signal Process. 57 (6), 2417–2420.

Long, T., Cohen, I., Berdugo, B., Yang, Y., Chen, J., 2019. Window-based constant beamwidth beamformer. Spec. Issue Sensors Speech Acoust. Audio Signal Process. Appl. Sens. 19, 1–20.

Moore, A.H., Evers, C., Naylor, P.A., 2017. Direction of arrival estimation in the spherical harmonic domain using subspace pseudointensity vectors. IEEE/ACM Trans. Audio Speech Lang. Process. 25 (1), 178–192.

Parra, L.C., 2006. Steerable frequency-invariant beamforming for arbitrary arrays. J. Acoust. Soc. Am. 119 (6), 3839–3847.

Pierce, A., 2019. Acoustics: An Introduction to its Physical Principles and Applications. Springer International Publishing, Switzerland.

Rafaely, B., 2015. Fundamentals of Spherical Array Processing. Springer-Verlag Berlin Heidelberg, Berlin.

Ribeiro, L.N., de Almeida, A.L.F., Mota, J.C.M., 2016. Tensor beamforming for multilinear translation invariant arrays. In: 2016 IEEE International Conference on Acoustics, Speech and Signal Processing. ICASSP, pp. 2966–2970.

Rix, A.W., Beerends, J.G., Hollier, M.P., Hekstra, A.P., 2001. 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, Vol. 2. pp. 749–752.

Rosen, O., Cohen, I., Malah, D., 2017. FIR-based symmetrical acoustic beamformer with a constant beamwidth. Signal Process. 130, 365–376.

Taal, C.H., Hendriks, R.C., Heusdens, R., Jensen, J., 2011. An algorithm for intelligibility prediction of time–frequency weighted noisy speech. IEEE Trans. Audio Speech Lang. Process. 19 (7), 2125–2136.

Tourbabin, V., Agmon, M., Rafaely, B., Tabrikian, J., 2012. Optimal real-weighted beamforming with application to linear and spherical arrays. IEEE Trans. Audio Speech Lang. Process. 20, 2575–2585.

Van Trees, H., 2004. Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory. In: Detection, Estimation, and Modulation Theory, Wiley, New York.

Wang, X., Benesty, J., Chen, J., Huang, G., Cohen, I., 2021. Beamforming with cube microphone arrays via Kronecker product decompositions. IEEE/ACM Trans. Audio Speech Lang. Process. 29, 1774–1784.

Wang, X., Huang, G., Cohen, I., Benesty, J., Chen, J., 2021. Kronecker product adaptive beamforming for microphone arrays. In: 2021 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC). pp. 49–54.

Zhuang, B., Rohling, R., Abolmaesumi, P., 2019. Region-of-interest-based closed-loop beamforming for spinal ultrasound imaging. IEEE Trans. Ultrason. Ferroelectr. Freq. Control 66 (8), 1266–1280.

Zoltowski, M.D., Haardt, M., Mathews, C.P., 1996. Closed-form 2-D angle estimation with rectangular arrays in element space or beamspace via unitary ESPRIT. IEEE Trans. Signal Process. 44 (2), 316–328.