On the Design of Differential Kronecker Product Beamformers

Gal Itzhak, Jacob Benesty, and Israel Cohen

***Abstract*—In this paper, we present a generalized approach for differential microphone array (DMA) beamforming in the short-time Fourier transform (STFT) domain. We propose a multistage beamforming approach, which considers a Kronecker product (KP) decomposition of the global beamformer into two independent sub-beamformers. We derive differential KP beamformers according to different criteria and analyze their performances, which are tuned by three design parameters. These parameters allow a high beamforming design flexibility; in particular, non-differential or non-KP beamformers may be obtained as special cases. Depending on the selection of parameters, we demonstrate a preferable performance with the new approach with respect to the white noise gain and directivity factor measures. In addition, we consider the task of speech enhancement. We show that differential KP beamformers perform better than non-differential and non-KP beamformers in terms of the quality and intelligibility of their respective time-domain enhanced signals, particularly in moderately reverberant environments.**

***Index Terms*—Microphone arrays, uniform linear arrays, differential beamforming, optimal beamformer, Kronecker product decomposition.**

I. Introduction

Communication systems in the real world involving audio and speech signals are typically required to operate in the presence of undesirable background noise that might heavily corrupt their quality. Taking advantage of a multichannel structure, the notion of beamforming has been in the center of attention to cope with this problem, being addressed in numerous researches [1]–[4]. Traced back to almost a century ago [5], [6], differential microphone arrays (DMAs) have been proposed and optimized with the underlying principle of exploiting acoustic pressure differences among adjacent microphones [7]–[11]. This principle implies arrays of small sizes and frequency-invariant beampatterns [12]–[16].

Typically, in order to design high-order DMAs which were capable of obtaining a significant amount of noise reduction, a multistage approach was taken. That is, the operation of differentiating acoustic pressure observations was successively repeated, in analogy to high-order derivatives of analytic functions [17], [18]. This approach was implemented in the time domain. Unfortunately, it was highly susceptible to array mismatches and imperfections [19]–[21], making it less appealing under practical conditions. Consequently, DMA design in the short-time Fourier transform (STFT) domain was introduced, providing a robust framework that is based on a single stage with linear matrix operations. Despite its simplicity, it is still capable of satisfying spatial constraints while simultaneously minimizing residual noise, in either fixed or adaptive settings [22], [23]. Nevertheless, due to the simple single-stage structure and inherent linear nature, the noise reduction capabilities of DMAs are limited. Recently, the design of DMAs in the STFT-domain was generalized to a multistage structure [24]. This technique was shown to be effective to reduce diffuse noise and to handle reverberant environments, though a significant drawback was its white noise amplification.

DMAs in the STFT domain were thoroughly analyzed and adapted into many variations. One recent example is Kronecker product (KP) beamforming, in which a global beamformer is decomposed into a KP of independent sub-beamformers that may be individually designed and optimized [25]–[30]. The main advantage of KP beamformers is their great design flexibility. That is, each sub-beamformer may be optimized by a different criterion, yielding a global beamformer that is "optimized" according to all criteria. The relative sizes of the sub-beamformers set the trade-off for the optimization of the global beamformer.

In this paper, we present a differential KP beamforming approach in the STFT domain, which generalizes the approach in [24]. We propose a multistage approach, which considers a KP decomposition of the global beamformer into two independent sub-beamformers. We derive differential KP beamformers according to different criteria and analyze their performances, tuned by three design parameters. These parameters facilitate high design flexibility; in particular, non-differential or non-KP beamformers may be obtained as special cases. Depending on the selection of parameters, we demonstrate a preferable performance using the new approach with respect to the white noise gain (WNG) and directivity factor (DF) measures. This may turn important in practice when considering microphones whose self-noise is significant or scenarios of considerable reverberation. In addition, we consider the task of speech enhancement. We show that differential KP beamformers perform better than non-differential and non-KP beamformers in terms of the quality and intelligibility of their respective time-domain enhanced signals, particularly in moderately reverberant environments.

The rest of the paper is organized as follows. In Section II, we briefly review the multistage differential beamforming approach in the STFT domain. In Section III, we present

This research was supported by the Pazy Research Foundation and the ISF-NSFC joint research program (grant No. 2514/17).

G. Itzhak and I. Cohen are with Andrew and Erna Viterby Faculty of Electrical Engineering, Technion–Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: galitz@campus.technion.ac.il, icohen@ee.technion.ac.il).

J. Benesty is with INRS-EMT, University of Quebec, 800 de la Gauchetiere Ouest, Montreal, QC H5A 1K6, Canada (e-mail: Jacob.Benesty@inrs.ca).

the differential KP beamforming approach and reformulate the problem in terms of two independent sub-beamformers. Then, in Section IV, performance measures are derived accordingly. We present the familiar beamforming design measures as they are expressed in the context of differential KP beamforming. Section V is dedicated to the derivation of five differential KP beamformers, each of which is designed with respect to a different optimization criterion. Then, in Section VI, we perform a simulative study in an anechoic environment, followed by simulations of speech signals in a reverberant room as well as with array imperfections. Finally, we summarize this work in Section VII.

II. Signal Model and problem Formulation

We consider a uniform linear array (ULA), consisting of $M \geq 2$ omnidirectional microphones with an interelement spacing equal to δ. Let us assume that a farfield plane wave propagates from an azimuth angle θ in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, and impinges on this ULA. In this scenario, the corresponding steering vector (of length M) is [31]

$$\mathbf{d}_{\theta,M}(f) = \left[1 \quad e^{-\jmath 2\pi f \delta \cos\theta/c} \quad \cdots \quad e^{-\jmath(M-1)2\pi f\delta\cos\theta/c}\right]^T, \quad (1)$$

where f is the temporal frequency, $\jmath = \sqrt{-1}$ is the imaginary unit, and the superscript ^{T} is the transpose operator.

In order to be in the optimal working conditions of differential beamforming, we assume that the desired source comes from the direction $\theta_{\mathrm{s}} = 0$ and δ is small [13]. Note that in Section VI-C we demonstrate that our proposed approach is robust to small deviations in the values of θ_{s} and δ. In this case, we can express the frequency-domain observed signal vector of length M as [32]

$$\begin{aligned}\mathbf{y}(f) &= \left[\, Y_1(f) \quad Y_2(f) \quad \cdots \quad Y_M(f) \,\right]^T \\ &= \mathbf{x}(f) + \mathbf{v}(f) \\ &= \mathbf{d}_{0,M}(f) X(f) + \mathbf{v}(f), \end{aligned} \quad (2)$$

where $Y_m(f)$ is the mth microphone signal, $\mathbf{x}(f) = \mathbf{d}_{0,M}(f) X(f)$, $\mathbf{d}_{0,M}(f)$ is the steering vector at $\theta = 0$, $X(f)$ is the zero-mean desired source signal, $\mathbf{v}(f)$ is the zero-mean additive noise signal vector defined similarly to $\mathbf{y}(f)$, and $X(f)$ and $\mathbf{v}(f)$ are incoherent. In the rest, in order to simplify the notation, we drop the dependence on the temporal frequency, f. For a small and compact array, it is reasonable to assume that the variance of the noise is the same at all sensors, i.e., $\phi_V = \phi_{V_1} = \phi_{V_2} = \cdots = \phi_{V_M}$, with $\phi_{V_m} = E\left(|V_m|^2\right)$, $m = 1, 2, \ldots, M$ and $E(\cdot)$ denoting mathematical expectation. The meaning of this assumption is as follows. Considering the self-noise of the microphones, it implies that all of them are, roughly, of the same kind and have the same level of imperfections. Considering directional interferences and spherically isotropic (diffuse) noise, it implies that the power of the received signals is similar in all microphones, that is, the distance between the signal source to the array reference microphone is much bigger than the interelement spacing. Clearly, this is a restatement of the farfield signal model. Consequently, the variance of Y_m is $\phi_{Y_m} = \phi_Y = \phi_X + \phi_V$, where ϕ_X is the variance of X.

Let P be a positive integer with $0 \leq P < M$. We can transform the observed signal vector $\mathbf{y}$ of length M to a Pth-order forward spatial difference of $\mathbf{y}$ of length $M(P) = M - P$, i.e., [24]

$$\mathbf{y}_{(P)} = \mathbf{\Delta}_{(P)}\mathbf{y}, \quad (3)$$

with $\mathbf{y}_{(0)} = \mathbf{y}$, where

$$\mathbf{\Delta}_{(P)} = \begin{bmatrix} \mathbf{c}_{(P)}^T & 0 & \cdots & 0 \\ 0 & \mathbf{c}_{(P)}^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{c}_{(P)}^T \end{bmatrix} \quad (4)$$

is a matrix of size $M(P) \times M$, with $\mathbf{\Delta}_{(0)} = \mathbf{I}_M$, which is the $M \times M$ identity matrix,

$$\mathbf{c}_{(P)} = \left[(-1)^P \binom{P}{0} \quad (-1)^{P-1}\binom{P}{1} \quad \cdots \quad -\binom{P}{P-1} \quad 1\right]^T \quad (5)$$

is a vector of length $P+1$, and

$$\binom{P}{j} = \frac{P!}{j!\,(P-j)!}$$

is the binomial coefficient. The major benefit with the difference observed signal vector, $\mathbf{y}_{(P)}$ of length $M(P)$, for $P > 0$, is that it is less sensitive to diffuse noise as compared to $\mathbf{y}$; in fact, the larger is the value of P, the higher is the signal-to-noise ratio (SNR) of $\mathbf{y}_{(P)}$. However, $\mathbf{y}_{(P)}$ is more sensitive to white noise.

It can be shown that (3) can be expressed as [24]

$$\begin{aligned}\mathbf{y}_{(P)} &= \tau_0^P \mathbf{d}_{0,M(P)} X + \mathbf{v}_{(P)} \\ &= \mathbf{x}_{(P)} + \mathbf{v}_{(P)}, \end{aligned} \quad (6)$$

where

$$\tau_0 = e^{-\jmath 2\pi f\delta/c} - 1 \quad (7)$$

is a frequency-dependent variable, $\mathbf{d}_{0,M(P)}$ is the steering vector of length $M(P)$ at $\theta = 0$, and $\mathbf{v}_{(P)} = \mathbf{\Delta}_{(P)}\mathbf{v}$. We deduce that the $M(P) \times M(P)$ covariance matrix of $\mathbf{y}_{(P)}$ is

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{y}_{(P)}} &= \phi_X |\tau_0|^{2P} \mathbf{d}_{0,M(P)}\mathbf{d}_{0,M(P)}^H + \mathbf{\Delta}_{(P)}\mathbf{\Phi}_{\mathbf{v}}\mathbf{\Delta}_{(P)}^T \\ &= \phi_X |\tau_0|^{2P} \mathbf{d}_{0,M(P)}\mathbf{d}_{0,M(P)}^H + \mathbf{\Phi}_{\mathbf{v}_{(P)}} \\ &= \phi_X |\tau_0|^{2P} \mathbf{d}_{0,M(P)}\mathbf{d}_{0,M(P)}^H + \phi_V \mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathbf{v}}\mathbf{\Delta}_{(P)}^T \\ &= \phi_X |\tau_0|^{2P} \mathbf{d}_{0,M(P)}\mathbf{d}_{0,M(P)}^H + \phi_V \mathbf{\Gamma}_{\mathbf{v}_{(P)}}, \end{aligned} \quad (8)$$

where $\mathbf{\Phi}_{\mathbf{v}}$ is the covariance matrix of $\mathbf{v}$, $\mathbf{\Gamma}_{\mathbf{v}} = \mathbf{\Phi}_{\mathbf{v}}/\phi_V$, and $\mathbf{\Gamma}_{\mathbf{v}_{(P)}} = \mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathbf{v}}\mathbf{\Delta}_{(P)}^T$.

III. Differential Kronecker Product Beamforming

Assume that $M(P) = M - P = M_{\mathbf{a}} \times M_{\mathbf{b}}$, where $M_{\mathbf{a}}, M_{\mathbf{b}} \geq 1$. Then, one can verify that the steering vector $\mathbf{d}_{\theta,M(P)}$ can be decomposed as [27]

$$\mathbf{d}_{\theta,M(P)} = \mathbf{a}_\theta \otimes \mathbf{b}_\theta, \tag{9}$$

where

$$\mathbf{a}_\theta = \Big[1 \;\; e^{-\jmath 2\pi f M_{\mathbf{b}} \delta \cos\theta/c} \;\; \cdots \;\; e^{-\jmath (M_{\mathbf{a}}-1) 2\pi f M_{\mathbf{b}} \delta \cos\theta/c} \;\Big]^T \tag{10}$$

is the steering vector (of length $M_{\mathbf{a}}$) corresponding to a ULA of $M_{\mathbf{a}}$ sensors with an interelement spacing equal to $M_{\mathbf{b}}\delta$, $\otimes$ is the Kronecker product, and

$$\mathbf{b}_\theta = \Big[1 \;\; e^{-\jmath 2\pi f \delta \cos\theta/c} \;\; \cdots \;\; e^{-\jmath (M_{\mathbf{b}}-1) 2\pi f \delta \cos\theta/c} \;\Big]^T \tag{11}$$

is the steering vector (of length $M_{\mathbf{b}}$) corresponding to a ULA of $M_{\mathbf{b}}$ sensors with an interelement spacing equal to δ. As a consequence, the signal model in (6) becomes

$$\mathbf{y}_{(P)} = \tau_0^P \left(\mathbf{a}_0 \otimes \mathbf{b}_0\right) X + \mathbf{v}_{(P)} \tag{12}$$

and its covariance matrix is

$$\mathbf{\Phi}_{\mathbf{y}_{(P)}} = \phi_X \left|\tau_0\right|^{2P} \left(\mathbf{a}_0\mathbf{a}_0^H\right) \otimes \left(\mathbf{b}_0\mathbf{b}_0^H\right) + \phi_V \mathbf{\Gamma}_{\mathbf{v}_{(P)}}. \tag{13}$$

Because of the particular structure of the steering vector in (9) and in order to fully exploit this structure, we propose (global) beamformers of the form:

$$\mathbf{h} = \mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}}, \tag{14}$$

where $\mathbf{h}_{\mathbf{a}}$ and $\mathbf{h}_{\mathbf{b}}$ are two complex-valued linear filters of lengths $M_{\mathbf{a}}$ and $M_{\mathbf{b}}$, respectively. Then, in the proposed context, linear beamforming is performed by applying $\mathbf{h}$ [from (14)] to $\mathbf{y}_{(P)}$ [from (12)], i.e.,

$$\begin{aligned} Z &= \mathbf{h}^H \mathbf{y}_{(P)} \\ &= \mathbf{h}^H \mathbf{x}_{(P)} + \mathbf{h}^H \mathbf{v}_{(P)} \\ &= X_{\mathrm{fd}} + V_{\mathrm{rn}}, \end{aligned} \tag{15}$$

where Z is the estimate of the desired signal, X,

$$\begin{aligned} X_{\mathrm{fd}} &= \tau_0^P \left(\mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}}\right)^H \left(\mathbf{a}_0 \otimes \mathbf{b}_0\right) X \\ &= \tau_0^P \left(\mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0\right) \left(\mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0\right) X \end{aligned} \tag{16}$$

is the filtered desired signal, and

$$V_{\mathrm{rn}} = \left(\mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}}\right)^H \mathbf{v}_{(P)} \tag{17}$$

is the residual noise. We deduce that the variance of Z is

$$\begin{aligned} \phi_Z = {} & \left|\tau_0\right|^{2P} \left|\mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0\right|^2 \left|\mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0\right|^2 \phi_X \\ & + \phi_V \left(\mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}}\right)^H \mathbf{\Gamma}_{\mathbf{v}_{(P)}} \left(\mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}}\right). \end{aligned} \tag{18}$$

We see from X_{fd} that the distortionless constraint is

$$\left(\mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0\right) \left(\mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0\right) = \tau_0^{-P}. \tag{19}$$

In the rest, we choose $\mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0 = \tau_0^{-P}$ and $\mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0 = 1$, so that (19) is satisfied.

Furthermore, we will often use the following relationships:

$$\begin{aligned} \mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}} &= \left(\mathbf{h}_{\mathbf{a}} \otimes \mathbf{I}_{M_{\mathbf{b}}}\right) \mathbf{h}_{\mathbf{b}} && (20) \\ &= \left(\mathbf{I}_{M_{\mathbf{a}}} \otimes \mathbf{h}_{\mathbf{b}}\right) \mathbf{h}_{\mathbf{a}}, && (21) \end{aligned}$$

where $\mathbf{I}_{M_{\mathbf{b}}}$ and $\mathbf{I}_{M_{\mathbf{a}}}$ are the identity matrices of sizes $M_{\mathbf{b}} \times M_{\mathbf{b}}$ and $M_{\mathbf{a}} \times M_{\mathbf{a}}$, respectively.

IV. Performance Measures

In this section, we express common performance measures according to the differential KP beamforming approach.

The first useful measure discussed in this section is the beampattern, which describes the sensitivity of the beamformer to a plane wave (source signal) impinging on the ULA from the direction θ. Mathematically, it is defined as

$$\begin{aligned} \mathcal{B}_\theta\left(\mathbf{h}\right) &= \tau_\theta^P \mathbf{h}^H \mathbf{d}_{\theta,M(P)} \\ &= \tau_\theta^P \times \mathbf{h}_{\mathbf{a}}^H \mathbf{a}_\theta \times \mathbf{h}_{\mathbf{b}}^H \mathbf{b}_\theta \\ &= \mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h}_{\mathbf{a}}\right) \times \mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h}_{\mathbf{b}}\right), \end{aligned} \tag{22}$$

where

$$\tau_\theta = e^{-\jmath 2\pi f \delta \cos\theta/c} - 1, \tag{23}$$

$\mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h}_{\mathbf{a}}\right) = \tau_\theta^P \mathbf{h}_{\mathbf{a}}^H \mathbf{a}_\theta$, and $\mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h}_{\mathbf{b}}\right) = \mathbf{h}_{\mathbf{b}}^H \mathbf{b}_\theta$. The global beampattern is composed of three terms: the first one, τ_θ^P, emphasizes the directivity of the pattern; the second term, $\mathbf{h}_{\mathbf{a}}^H \mathbf{a}_\theta$, is the beampattern of the first ULA with an interelement spacing equal to $M_{\mathbf{b}}\delta$, and the last term, $\mathbf{h}_{\mathbf{b}}^H \mathbf{b}_\theta$, corresponds to the beampattern of the second ULA with an interelement spacing equal to δ.

From (2), we easily find that the input SNR is

$$\mathrm{iSNR} = \frac{\phi_X}{\phi_V}. \tag{24}$$

The output SNR is defined [from the variance of Z, see (18)] as

$$\mathrm{oSNR}\left(\mathbf{h}\right) = \frac{\phi_X}{\phi_V} \times \frac{\left|\tau_0\right|^{2P} \left|\mathbf{h}^H \mathbf{d}_{\theta,M(P)}\right|^2}{\mathbf{h}^H \mathbf{\Gamma}_{\mathbf{v}_{(P)}} \mathbf{h}}. \tag{25}$$

The definition of the gain in SNR is obtained from the previous definitions, i.e.,

$$\begin{aligned} \mathcal{G}\left(\mathbf{h}\right) &= \frac{\mathrm{oSNR}\left(\mathbf{h}\right)}{\mathrm{iSNR}} \\ &= \frac{\left|\tau_0\right|^{2P} \left|\mathbf{h}^H \mathbf{d}_{\theta,M(P)}\right|^2}{\mathbf{h}^H \mathbf{\Gamma}_{\mathbf{v}_{(P)}} \mathbf{h}}. \end{aligned} \tag{26}$$

We can rewrite this gain as

$$\mathcal{G}\left(\mathbf{h}_{\mathbf{a}}, \mathbf{h}_{\mathbf{b}}\right) = \frac{\left|\tau_0\right|^{2P} \left|\mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0\right|^2 \left|\mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0\right|^2}{\left(\mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}}\right)^H \mathbf{\Gamma}_{\mathbf{v}_{(P)}} \left(\mathbf{h}_{\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b}}\right)}. \tag{27}$$

When $\mathbf{h}_{\mathbf{b}}$ is fixed, given, and satisfies the distortionless constraint, i.e., $\mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0 = 1$; then, we can express the gain as

$$\mathcal{G}\left(\mathbf{h}_{\mathbf{a}} \left| \mathbf{h}_{\mathbf{b}}\right.\right) = \frac{\left|\tau_0\right|^{2P} \left|\mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0\right|^2}{\mathbf{h}_{\mathbf{a}}^H \mathbf{\Gamma}_{\mathbf{v}_{(P)},\mathbf{b}} \mathbf{h}_{\mathbf{a}}}, \tag{28}$$

where

$$\mathbf{\Gamma}_{\mathbf{v}_{(P)},\mathbf{b}} = \left(\mathbf{I}_{M_{\mathbf{a}}} \otimes \mathbf{h}_{\mathbf{b}}\right)^H \mathbf{\Gamma}_{\mathbf{v}_{(P)}} \left(\mathbf{I}_{M_{\mathbf{a}}} \otimes \mathbf{h}_{\mathbf{b}}\right). \tag{29}$$

In the same way, when $\mathbf{h_a}$ is fixed, given, and satisfies the distortionless constraint, i.e., $\mathbf{h_a}^H\mathbf{a}_0 = \tau_0^{-P}$; then, we can express the gain as

$$\mathcal{G}\left(\mathbf{h_b}\left|\mathbf{h_a}\right.\right) = \frac{\left|\mathbf{h_b}^H\mathbf{b}_0\right|^2}{\mathbf{h_b}^H\mathbf{\Gamma}_{\mathbf{v}_{(P)},\mathbf{a}}\mathbf{h_b}}, \tag{30}$$

where

$$\mathbf{\Gamma}_{\mathbf{v}_{(P)},\mathbf{a}} = \left(\mathbf{h_a}\otimes\mathbf{I}_{M_\mathbf{b}}\right)^H\mathbf{\Gamma}_{\mathbf{v}_{(P)}}\left(\mathbf{h_a}\otimes\mathbf{I}_{M_\mathbf{b}}\right). \tag{31}$$

One important measure which expresses the sensitivity of the ULA to spatially white noise is the WNG, which is defined by taking $\mathbf{\Gamma_v} = \mathbf{I}_M$, i.e.,

$$\begin{aligned}\mathcal{W}\left(\mathbf{h}\right) &= \frac{\left|\tau_0\right|^{2P}\left|\mathbf{h}^H\mathbf{d}_{\theta,M(P)}\right|^2}{\mathbf{h}^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\mathbf{h}}\\ &= \frac{\left|\tau_0\right|^{2P}\left|\mathbf{h_a}^H\mathbf{a}_0\right|^2\left|\mathbf{h_b}^H\mathbf{b}_0\right|^2}{\left(\mathbf{h_a}\otimes\mathbf{h_b}\right)^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\left(\mathbf{h_a}\otimes\mathbf{h_b}\right)}\\ &= \mathcal{W}\left(\mathbf{h_a},\mathbf{h_b}\right),\end{aligned} \tag{32}$$

which can only be expressed as a product of the WNGs of the two sub-beamformers for $P = 0$.

When $\mathbf{h_b}$ is fixed, given, and satisfies the distortionless constraint, i.e., $\mathbf{h_b}^H\mathbf{b}_0 = 1$; then, we can express the WNG as

$$\mathcal{W}\left(\mathbf{h_a}\left|\mathbf{h_b}\right.\right) = \frac{\left|\tau_0\right|^{2P}\left|\mathbf{h_a}^H\mathbf{a}_0\right|^2}{\mathbf{h_a}^H\left[\left(\mathbf{I}_{M_\mathbf{a}}\otimes\mathbf{h_b}\right)^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\left(\mathbf{I}_{M_\mathbf{a}}\otimes\mathbf{h_b}\right)\right]\mathbf{h_a}}. \tag{33}$$

Similarly, when $\mathbf{h_a}$ is fixed, given, and satisfies the distortionless constraint, i.e., $\mathbf{h_a}^H\mathbf{a}_0 = \tau_0^{-P}$ we have

$$\mathcal{W}\left(\mathbf{h_b}\left|\mathbf{h_a}\right.\right) = \frac{\left|\mathbf{h_b}^H\mathbf{b}_0\right|^2}{\mathbf{h_b}^H\left[\left(\mathbf{h_a}\otimes\mathbf{I}_{M_\mathbf{b}}\right)^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\left(\mathbf{h_a}\otimes\mathbf{I}_{M_\mathbf{b}}\right)\right]\mathbf{h_b}}. \tag{34}$$

Another important measure, which quantifies how the microphone array performs in the presence of reverberation is the DF. Considering the spherically isotropic (diffuse) noise field, the DF is defined as

$$\begin{aligned}\mathcal{D}\left(\mathbf{h}\right) &= \frac{\left|\mathcal{B}_0\left(\mathbf{h}\right)\right|^2}{\frac{1}{2}\int_0^\pi\left|\mathcal{B}_\theta\left(\mathbf{h}\right)\right|^2\sin\theta d\theta}\\ &= \frac{\left|\mathcal{B}_{0,\mathbf{a}}\left(\mathbf{h_a}\right)\right|^2\left|\mathcal{B}_{0,\mathbf{b}}\left(\mathbf{h_b}\right)\right|^2}{\frac{1}{2}\int_0^\pi\left|\mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h_a}\right)\right|^2\left|\mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h_b}\right)\right|^2\sin\theta d\theta}\\ &= \frac{\left|\tau_0\right|^{2P}\left|\mathbf{h_a}^H\mathbf{a}_0\right|^2\left|\mathbf{h_b}^H\mathbf{b}_0\right|^2}{\left(\mathbf{h_a}\otimes\mathbf{h_b}\right)^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\left(\mathbf{h_a}\otimes\mathbf{h_b}\right)}\\ &= \mathcal{D}\left(\mathbf{h_a},\mathbf{h_b}\right),\end{aligned} \tag{35}$$

where the elements of the diffuse noise coherence matrix $\mathbf{\Gamma}_{\mathrm{d}}$ are given by

$$\left(\mathbf{\Gamma}_{\mathrm{d}}\right)_{ij} = \frac{\sin\left[2\pi f(j-i)\delta/c\right]}{2\pi f(j-i)\delta/c}. \tag{36}$$

When $\mathbf{h_b}$ is fixed, given, and satisfies the distortionless constraint, i.e., $\mathbf{h_b}^H\mathbf{b}_0 = 1$; then, we can express the DF as

$$\mathcal{D}\left(\mathbf{h_a}\left|\mathbf{h_b}\right.\right) = \frac{\left|\tau_0\right|^{2P}\left|\mathbf{h_a}^H\mathbf{a}_0\right|^2}{\mathbf{h_a}^H\mathbf{\Gamma}_{\mathrm{d},\mathbf{b}}\mathbf{h_a}}, \tag{37}$$

where

$$\mathbf{\Gamma}_{\mathrm{d},\mathbf{b}} = \frac{1}{2}\int_0^\pi\left|\tau_\theta\right|^{2P}\mathbf{a}_\theta\mathbf{a}_\theta^H\left|\mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h_b}\right)\right|^2\sin\theta d\theta. \tag{38}$$

In an analogous way, when $\mathbf{h_a}$ is fixed, given, and satisfies the distortionless constraint, i.e., $\mathbf{h_a}^H\mathbf{a}_0 = \tau_0^{-P}$; then, we can write the DF as

$$\mathcal{D}\left(\mathbf{h_b}\left|\mathbf{h_a}\right.\right) = \frac{\left|\mathbf{h_b}^H\mathbf{b}_0\right|^2}{\mathbf{h_b}^H\mathbf{\Gamma}_{\mathrm{d},\mathbf{a}}\mathbf{h_b}}, \tag{39}$$

where

$$\mathbf{\Gamma}_{\mathrm{d},\mathbf{a}} = \frac{1}{2}\int_0^\pi\mathbf{b}_\theta\mathbf{b}_\theta^H\left|\mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h_a}\right)\right|^2\sin\theta d\theta. \tag{40}$$

Finally, the last measure of interest in this section is the front-to-back ratio (FBR), which is defined as the ratio of the power of the output of the array to signals propagating from the front-half plane to the output power for signals arriving from the rear-half plane. This ratio, for the spherically isotropic (diffuse) noise field, is mathematically defined as

$$\begin{aligned}\mathcal{F}\left(\mathbf{h}\right) &= \frac{\int_0^{\pi/2}\left|\mathcal{B}_\theta\left(\mathbf{h}\right)\right|^2\sin\theta d\theta}{\int_{\pi/2}^{\pi}\left|\mathcal{B}_\theta\left(\mathbf{h}\right)\right|^2\sin\theta d\theta}\\ &= \frac{\int_0^{\pi/2}\left|\mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h_a}\right)\right|^2\left|\mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h_b}\right)\right|^2\sin\theta d\theta}{\int_{\pi/2}^{\pi}\left|\mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h_a}\right)\right|^2\left|\mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h_b}\right)\right|^2\sin\theta d\theta}\\ &= \mathcal{F}\left(\mathbf{h_a},\mathbf{h_b}\right).\end{aligned} \tag{41}$$

When $\mathbf{h_b}$ is fixed and given; then, we can express the FBR as

$$\mathcal{F}\left(\mathbf{h_a}\left|\mathbf{h_b}\right.\right) = \frac{\mathbf{h_a}^H\mathbf{\Gamma}_{\mathrm{f},\mathbf{b}}\mathbf{h_a}}{\mathbf{h_a}^H\mathbf{\Gamma}_{\mathrm{b},\mathbf{b}}\mathbf{h_a}}, \tag{42}$$

where

$$\mathbf{\Gamma}_{\mathrm{f},\mathbf{b}} = \int_0^{\pi/2}\left|\tau_\theta\right|^{2P}\mathbf{a}_\theta\mathbf{a}_\theta^H\left|\mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h_b}\right)\right|^2\sin\theta d\theta, \tag{43}$$

$$\mathbf{\Gamma}_{\mathrm{b},\mathbf{b}} = \int_{\pi/2}^{\pi}\left|\tau_\theta\right|^{2P}\mathbf{a}_\theta\mathbf{a}_\theta^H\left|\mathcal{B}_{\theta,\mathbf{b}}\left(\mathbf{h_b}\right)\right|^2\sin\theta d\theta. \tag{44}$$

Similarly, when $\mathbf{h_a}$ is fixed and given; then, we can express the FBR as

$$\mathcal{F}\left(\mathbf{h_b}\left|\mathbf{h_a}\right.\right) = \frac{\mathbf{h_b}^H\mathbf{\Gamma}_{\mathrm{f},\mathbf{a}}\mathbf{h_b}}{\mathbf{h_b}^H\mathbf{\Gamma}_{\mathrm{b},\mathbf{a}}\mathbf{h_b}}, \tag{45}$$

where

$$\mathbf{\Gamma}_{\mathrm{f},\mathbf{a}} = \int_0^{\pi/2}\mathbf{b}_\theta\mathbf{b}_\theta^H\left|\mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h_a}\right)\right|^2\sin\theta d\theta, \tag{46}$$

$$\mathbf{\Gamma}_{\mathrm{b},\mathbf{a}} = \int_{\pi/2}^{\pi}\mathbf{b}_\theta\mathbf{b}_\theta^H\left|\mathcal{B}_{\theta,\mathbf{a}}\left(\mathbf{h_a}\right)\right|^2\sin\theta d\theta. \tag{47}$$

V. Examples of Optimal Differential KP Beamformers

In this section, we propose five types of differential KP beamformers, each designed with respect to a different optimization criterion.

A. Maximum White Noise Gain

Let us start with the WNG measure. It is not possible to get a closed-form expression beamformer from the maximization of the WNG. However, using the alternating least-squares (ALS) strategy [33], [34], we can derive the maximum WNG (MWNG) beamformer iteratively from

$$\min_{\mathbf{h_a}} \mathbf{h}_{\mathbf{a}}^H \mathbf{\Gamma}_{\Delta,\mathbf{b}} \mathbf{h}_{\mathbf{a}} \quad \text{s. t.} \quad \mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0 = \tau_0^{-P}, \tag{48}$$

$$\min_{\mathbf{h_b}} \mathbf{h}_{\mathbf{b}}^H \mathbf{\Gamma}_{\Delta,\mathbf{a}} \mathbf{h}_{\mathbf{b}} \quad \text{s. t.} \quad \mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0 = 1, \tag{49}$$

whose solution is easily obtained by

$$\mathbf{h}_{\mathbf{a},\mathrm{MWNG}} = \frac{{\mathbf{\Gamma}_{\Delta,\mathbf{b}}}^{-1} \mathbf{a}_0}{\left(\tau_0^P\right)^* \mathbf{a}_0^H {\mathbf{\Gamma}_{\Delta,\mathbf{b}}}^{-1} \mathbf{a}_0}, \tag{50}$$

$$\mathbf{h}_{\mathbf{b},\mathrm{MWNG}} = \frac{{\mathbf{\Gamma}_{\Delta,\mathbf{a}}}^{-1} \mathbf{b}_0}{\mathbf{b}_0^H {\mathbf{\Gamma}_{\Delta,\mathbf{a}}}^{-1} \mathbf{b}_0}, \tag{51}$$

where

$$\mathbf{\Gamma}_{\Delta,\mathbf{a}} = (\mathbf{h}_{\mathbf{a},\mathrm{MWNG}} \otimes \mathbf{I}_{M_\mathbf{b}})^H \mathbf{\Delta}_{(P)} \mathbf{\Delta}_{(P)}^T (\mathbf{h}_{\mathbf{a},\mathrm{MWNG}} \otimes \mathbf{I}_{M_\mathbf{b}}), \tag{52}$$

$$\mathbf{\Gamma}_{\Delta,\mathbf{b}} = (\mathbf{I}_{M_\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MWNG}})^H \mathbf{\Delta}_{(P)} \mathbf{\Delta}_{(P)}^T (\mathbf{I}_{M_\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MWNG}}). \tag{53}$$

As a result, at iteration n, the global MWNG beamformer is

$$\mathbf{h}_{\mathrm{MWNG}}^{(n)} = \mathbf{h}_{\mathbf{a},\mathrm{MWNG}}^{(n)} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MWNG}}^{(n)}, \tag{54}$$

where

$$\mathbf{h}_{\mathbf{a},\mathrm{MWNG}}^{(n)} = \frac{{\mathbf{\Gamma}_{\Delta,\mathbf{b}}^{(n)}}^{-1} \mathbf{a}_0}{\left(\tau_0^P\right)^* \mathbf{a}_0^H {\mathbf{\Gamma}_{\Delta,\mathbf{b}}^{(n)}}^{-1} \mathbf{a}_0}, \tag{55}$$

$$\mathbf{h}_{\mathbf{b},\mathrm{MWNG}}^{(n)} = \frac{{\mathbf{\Gamma}_{\Delta,\mathbf{a}}^{(n)}}^{-1} \mathbf{b}_0}{\mathbf{b}_0^H {\mathbf{\Gamma}_{\Delta,\mathbf{a}}^{(n)}}^{-1} \mathbf{b}_0}, \tag{56}$$

and the iteratively updated coherence matrices are given by

$$\mathbf{\Gamma}_{\Delta,\mathbf{a}}^{(n)} = (\mathbf{h}_{\mathbf{a},\mathrm{MWNG}}^{(n)} \otimes \mathbf{I}_{M_\mathbf{b}})^H \mathbf{\Delta}_{(P)} \mathbf{\Delta}_{(P)}^T (\mathbf{h}_{\mathbf{a},\mathrm{MWNG}}^{(n)} \otimes \mathbf{I}_{M_\mathbf{b}}), \tag{57}$$

$$\mathbf{\Gamma}_{\Delta,\mathbf{b}}^{(n)} = (\mathbf{I}_{M_\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MWNG}}^{(n-1)})^H \mathbf{\Delta}_{(P)} \mathbf{\Delta}_{(P)}^T (\mathbf{I}_{M_\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MWNG}}^{(n-1)}). \tag{58}$$

B. Null Steering

Now, assume that we have one interference source impinging on the array from the direction $\theta_\mathrm{i} \neq \theta_\mathrm{s} = 0$ that we would like to completely cancel, i.e., to steer a null in that direction and, meanwhile, recover the desired source coming from the direction $\theta_\mathrm{s} = 0$. One possible approach to achieve such a behaviour is to have a null in the beampattern $\mathcal{B}_{\theta,\mathbf{b}}(\mathbf{h}_\mathbf{b})$, which implies a null in the global beampattern $\mathcal{B}_\theta(\mathbf{h})$, and a null constraint on the filter $\mathbf{h}_\mathbf{b}$. Then, by including the distortionless constraint, we can write the constraint equation as

$$\mathbf{C}^H \mathbf{h}_\mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{59}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{b}_0 & \mathbf{b}_{\theta_\mathrm{i}} \end{bmatrix} \tag{60}$$

is the constraint matrix of size $M_\mathbf{b} \times 2$ whose two columns are linearly independent. To find this filter, we maximize the WNG by taking (59) into account, i.e.,

$$\min_{\mathbf{h_b}} \mathbf{h}_{\mathbf{b}}^H \mathbf{\Gamma}_{\Delta,\mathbf{a}} \mathbf{h}_{\mathbf{b}} \quad \text{s. t.} \quad \mathbf{C}^H \mathbf{h}_\mathbf{b} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{61}$$

From this criterion, we get the following null-steering (NS) beamformer:

$$\mathbf{h}_{\mathbf{b},\mathrm{NS}} = \mathbf{\Gamma}_{\Delta,\mathbf{a}}^{-1} \mathbf{C} \left[\mathbf{C}^H \mathbf{\Gamma}_{\Delta,\mathbf{a}}^{-1} \mathbf{C}\right]^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{62}$$

For the other beamformer, we may choose $\mathbf{h}_{\mathbf{a},\mathrm{MWNG}}$ in (50). Therefore, the (global) proposed NS beamformer is obtained, iteratively, by

$$\mathbf{h}_{\mathrm{NS}}^{(n)} = \mathbf{h}_{\mathbf{a},\mathrm{MWNG}}^{(n)} \otimes \mathbf{h}_{\mathbf{b},\mathrm{NS}}^{(n)}, \tag{63}$$

where $\mathbf{h}_{\mathbf{a},\mathrm{MWNG}}^{(n)}$ is identical to the expression in (55) and

$$\mathbf{h}_{\mathbf{b},\mathrm{NS}}^{(n)} = {\mathbf{\Gamma}_{\Delta,\mathbf{a}}^{(n)}}^{-1} \mathbf{C} \left[\mathbf{C}^H {\mathbf{\Gamma}_{\Delta,\mathbf{a}}^{(n)}}^{-1} \mathbf{C}\right]^{-1} \begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{64}$$

C. Maximum Directivity Factor

Let us focus on the DF measure. In a similar manner to the MWNG beamformer, we can derive the maximum DF (MDF) beamformer iteratively from

$$\min_{\mathbf{h_a}} \mathbf{h}_{\mathbf{a}}^H \mathbf{\Gamma}_{\mathrm{d},\mathbf{b}} \mathbf{h}_{\mathbf{a}} \quad \text{s. t.} \quad \mathbf{h}_{\mathbf{a}}^H \mathbf{a}_0 = \tau_0^{-P}, \tag{65}$$

$$\min_{\mathbf{h_b}} \mathbf{h}_{\mathbf{b}}^H \mathbf{\Gamma}_{\mathrm{d},\mathbf{a}} \mathbf{h}_{\mathbf{b}} \quad \text{s. t.} \quad \mathbf{h}_{\mathbf{b}}^H \mathbf{b}_0 = 1. \tag{66}$$

We get

$$\mathbf{h}_{\mathbf{a},\mathrm{MDF}} = \frac{\mathbf{\Gamma}_{\mathrm{d},\mathbf{b}}^{-1} \mathbf{a}_0}{\left(\tau_0^P\right)^* \mathbf{a}_0^H \mathbf{\Gamma}_{\mathrm{d},\mathbf{b}}^{-1} \mathbf{a}_0}, \tag{67}$$

$$\mathbf{h}_{\mathbf{b},\mathrm{MDF}} = \frac{\mathbf{\Gamma}_{\mathrm{d},\mathbf{a}}^{-1} \mathbf{b}_0}{\mathbf{b}_0^H \mathbf{\Gamma}_{\mathrm{d},\mathbf{a}}^{-1} \mathbf{b}_0}. \tag{68}$$

As a result, at iteration n, the MDF beamformer is

$$\mathbf{h}_{\mathrm{MDF}}^{(n)} = \mathbf{h}_{\mathbf{a},\mathrm{MDF}}^{(n)} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MDF}}^{(n)}, \tag{69}$$

where

$$\mathbf{h}_{\mathbf{a},\mathrm{MDF}}^{(n)} = \frac{{\mathbf{\Gamma}_{\mathrm{d},\mathbf{b}}^{(n)}}^{-1} \mathbf{a}_0}{\left(\tau_0^P\right)^* \mathbf{a}_0^H {\mathbf{\Gamma}_{\mathrm{d},\mathbf{b}}^{(n)}}^{-1} \mathbf{a}_0}, \tag{70}$$

$$\mathbf{h}_{\mathbf{b},\mathrm{MDF}}^{(n)} = \frac{{\mathbf{\Gamma}_{\mathrm{d},\mathbf{a}}^{(n)}}^{-1} \mathbf{b}_0}{\mathbf{b}_0^H {\mathbf{\Gamma}_{\mathrm{d},\mathbf{a}}^{(n)}}^{-1} \mathbf{b}_0}, \tag{71}$$

and the iteratively updated coherence matrices are

$$\mathbf{\Gamma}_{\mathrm{d},\mathbf{a}}^{(n)} = \left(\mathbf{h}_{\mathbf{a},\mathrm{MDF}}^{(n)} \otimes \mathbf{I}_{M_\mathbf{b}}\right)^H \mathbf{\Gamma}_\mathrm{d} \left(\mathbf{h}_{\mathbf{a},\mathrm{MDF}}^{(n)} \otimes \mathbf{I}_{M_\mathbf{b}}\right), \quad (72)$$

$$\mathbf{\Gamma}_{\mathrm{d},\mathbf{b}}^{(n)} = \left(\mathbf{I}_{M_\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MDF}}^{(n-1)}\right)^H \mathbf{\Gamma}_\mathrm{d} \left(\mathbf{I}_{M_\mathbf{a}} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MDF}}^{(n-1)}\right). \quad (73)$$

We end this part by addressing a possible compromise between the WNG and DF measures. That is, taking advantage of the KP decomposition, we can combine the MWNG and MDF beamformers. For example, $\mathbf{h}_\mathbf{a}$ may be designed as a MWNG beamformer (50) whereas $\mathbf{h}_\mathbf{b}$ can be chosen as the MDF beamformer in (68). This would yield the global combined MWNG/MDF beamformer, which, at iteration n, is given by

$$\mathbf{h}_{\mathrm{MWNG/MDF}}^{(n)} = \mathbf{h}_{\mathbf{a},\mathrm{MWNG}}^{(n)} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MDF}}^{(n)}. \quad (74)$$

D. Maximum Front-to-Back Ratio

We turn our attention to the FBR measure. Employing the ALS strategy, the maximum FBR (MFBR) beamformer is derived from

$$\max_{\mathbf{h}_\mathbf{a}} \mathcal{F}\left(\mathbf{h}_\mathbf{a} \left| \mathbf{h}_\mathbf{b} \right.\right), \quad (75)$$

$$\max_{\mathbf{h}_\mathbf{b}} \mathcal{F}\left(\mathbf{h}_\mathbf{b} \left| \mathbf{h}_\mathbf{a} \right.\right). \quad (76)$$

Let $\mathbf{t}_\mathbf{a}$ (resp. $\mathbf{t}_\mathbf{b}$) be the eigenvector corresponding to the maximum eigenvalue of $\mathbf{\Gamma}_{\mathrm{b},\mathbf{b}}^{-1}\mathbf{\Gamma}_{\mathrm{f},\mathbf{b}}$ (resp. $\mathbf{\Gamma}_{\mathrm{b},\mathbf{a}}^{-1}\mathbf{\Gamma}_{\mathrm{f},\mathbf{a}}$). Then, the solutions are

$$\mathbf{h}_{\mathbf{a},\mathrm{MFBR}} = \frac{\mathbf{t}_\mathbf{a}}{\left(\tau_0^P\right)^* \mathbf{a}_0^H \mathbf{t}_\mathbf{a}}, \quad (77)$$

$$\mathbf{h}_{\mathbf{b},\mathrm{MFBR}} = \frac{\mathbf{t}_\mathbf{b}}{\mathbf{b}_0^H \mathbf{t}_\mathbf{b}}, \quad (78)$$

where we took into account the distortionless constraints. We deduce that, at iteration n, the MFBR beamformer is

$$\mathbf{h}_{\mathrm{MFBR}}^{(n)} = \mathbf{h}_{\mathbf{a},\mathrm{MFBR}}^{(n)} \otimes \mathbf{h}_{\mathbf{b},\mathrm{MFBR}}^{(n)}, \quad (79)$$

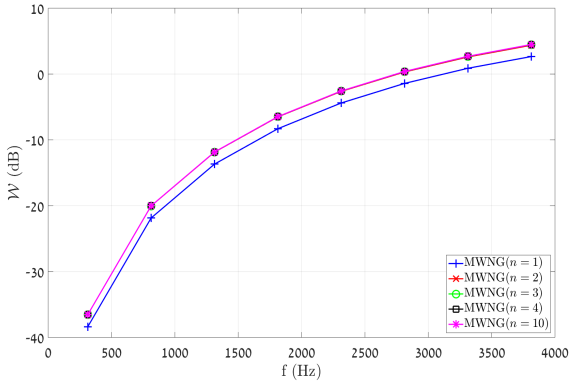
in which the appropriate coherence matrices are iteratively updated in an identical manner to the MWNG and MDF beamformers.

VI. Simulations

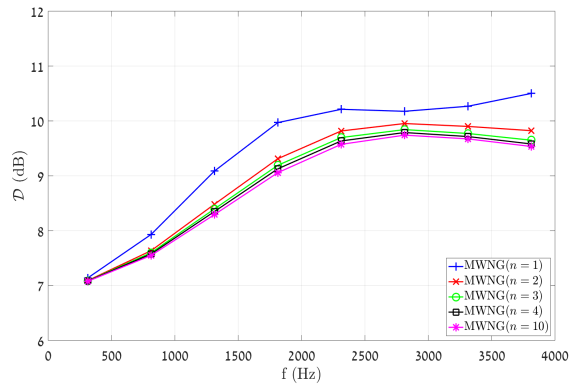
A. Performance Study

In this part, we investigate the performance of each of the differential KP beamformers presented in the former section in anechoic environments and with different settings.

Let us begin with the MWNG differential KP beamformer, $\mathbf{h}_\mathrm{MWNG}$. We note that as all other differential KP beamformers presented in this paper, it is derived iteratively. Therefore, it is important to get a notion of the convergence rate, i.e., to find a satisfactory value of the number of iterations n. Fig. 1 shows the WNG and DF measures with $(P = 2, M_\mathbf{a} = 4, M_\mathbf{b} = 2)$ for varying values of n and with initial sub-beamformers which implement the identity function: a one in the first element and zeros in the others. We observe that even for $n = 2$ convergence is achieved, whereas the performances with $n = 4$ and $n = 10$ nearly overlap. This result was verified with all other beamformers and with varying values of



(a)



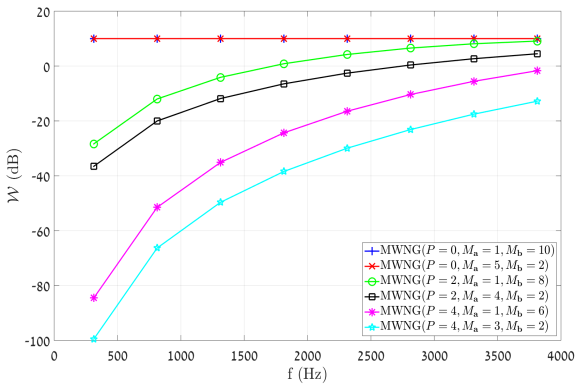
(b)

Fig. 1: WNG and DF measures with the MWNG differential KP beamformer, $\mathbf{h}_\mathrm{MWNG}$, with a varying number of iterations n. Simulation parameters: $M = 10$, $(P = 2, M_\mathbf{a} = 4, M_\mathbf{b} = 2)$ and $\delta = 1$ cm. (a) WNG and (b) DF.

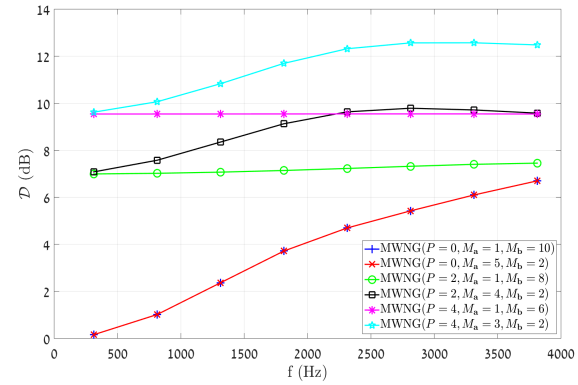
$(P, M_\mathbf{a}, M_\mathbf{b})$, to which we refer to from now on as the "array settings." Consequently, we will use the value of $n = 4$ in the rest.

Next, we explore the WNG and DF performance of $\mathbf{h}_\mathrm{MWNG}$ as a function of the array settings. These are demonstrated in Fig. 2. To begin with, we note that in the non-differential case ($P = 0$) the KP ($M_\mathbf{a} \neq 1$ and $M_\mathbf{b} \neq 1$) and non-KP ($M_\mathbf{a} = 1$ or $M_\mathbf{b} = 1$) settings result in similar beamformers whose performances perfectly overlap. This is with accordance to Section V-A as in this case $\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T = \mathbf{I}_M$, $\mathbf{h}_\mathrm{MWNG}$ converges to the MWNG beamformer of [28] and $\mathcal{W}\left(\mathbf{h}_\mathrm{MWNG}\right) = M$ regardless of $M_\mathbf{a}$ and $M_\mathbf{b}$. As P increases, the DF performance improves but the WNG deteriorates. In addition, for a fixed value of P, the influence of $M_\mathbf{a}$ and $M_\mathbf{b}$ is clearly observed. That is, with the non-KP settings, a preferable WNG performance is attained, whereas with the KP settings, the DF performance is preferable. Nonetheless, we observe, for example, that with $(2, 4, 2)$ both measures are higher than with $(4, 1, 6)$ for frequencies above 2500 Hz. We infer that the parameter P has a more dominant influence on the WNG-DF performance trade-off, whereas $M_\mathbf{a}$ and $M_\mathbf{b}$ enable a more flexible, finer, tuning.

Next, we analyze the NS differential KP beamformer, $\mathbf{h}_\mathrm{NS}$.

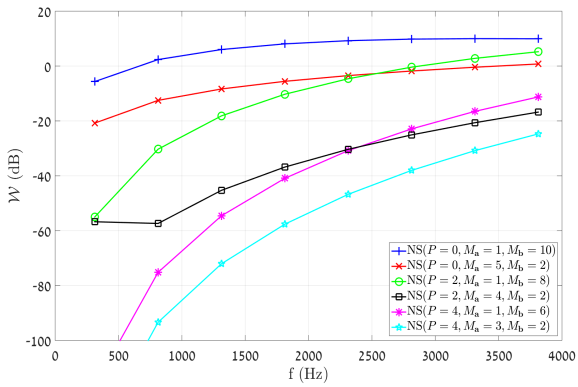


(a)

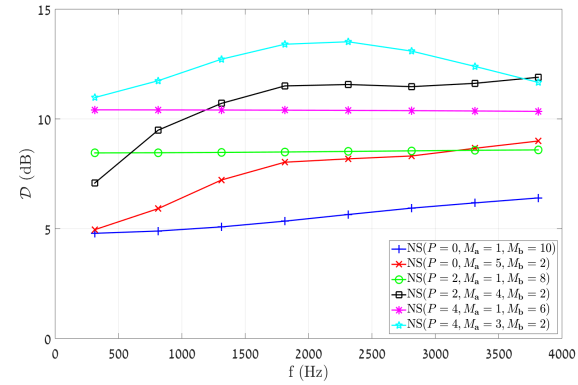


(b)

Fig. 2: WNG and DF measures with the MWNG differential KP beamformer, $\mathbf{h}_{\mathrm{MWNG}}$, with different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation parameters: $M = 10$ and $\delta = 1$ cm. (a) WNG and (b) DF.



(a)



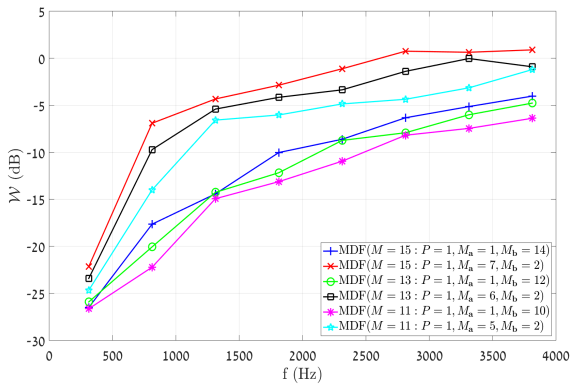
(b)

Fig. 3: WNG and DF measures with the NS differential KP beamformer, $\mathbf{h}_{\mathrm{NS}}$, with different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation parameters: $M = 10$ and $\delta = 1$ cm. (a) WNG and (b) DF.

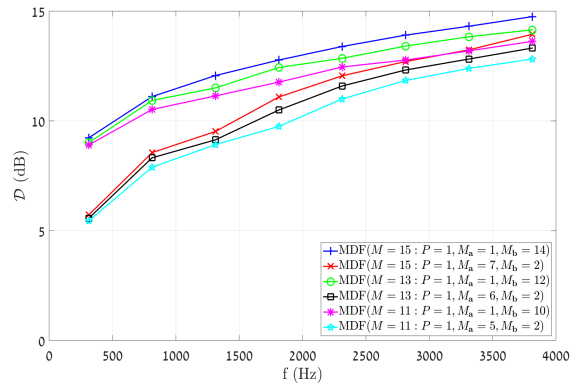
In our scenarios, the MN beamformer $\mathbf{h}_{\mathbf{b},\mathrm{MN}}$ is designed with a unique null constraint in the direction $\theta_{\mathrm{i}} = 90^{o}$. The corresponding WNG and DF measures are depicted in Fig. 3. To begin with, we observe that in contrast to the MWNG differential KP beamformer, with the NS differential KP beamformer the values of $M_{\mathbf{a}}$ and $M_{\mathbf{b}}$ set the performance trade-off for a fixed value of P even with $P = 0$. As with $\mathbf{h}_{\mathrm{MWNG}}$, the larger P the better the DF at the expense of the WNG, whereas the values of $M_{\mathbf{a}}$ and $M_{\mathbf{b}}$ tune the trade-off even further. That is, with the KP settings the DF performance is preferable with respect to the non-KP settings but the WNG is worse. Additionally, comparing the WNG performance of $\mathbf{h}_{\mathrm{MWNG}}$ and $\mathbf{h}_{\mathrm{NS}}$ for fixed settings, we note that the latter is lower as a consequence of the additional null constraint.

We now focus on the MDF differential KP beamformer, $\mathbf{h}_{\mathrm{MDF}}$, and the MWNG/MDF differential KP beamformer, $\mathbf{h}_{\mathrm{MWNG/MDF}}$, whose performances with arrays of $M = 11$, 13, and 15 microphones are shown in Fig. 4 and Fig. 5, respectively. As for high values of P white noise amplification was shown to be significant, we set $P = 1$ and examine the performance differences between the KP and non-KP versions of the differential beamformers. We point out that we use a small regularization factor of $\lambda = 10^{-2}$ when inverting the diffuse noise coherence matrix. Firstly, we note that with the non-KP settings $\mathbf{h}_{\mathrm{MDF}}$ and $\mathbf{h}_{\mathrm{MWNG/MDF}}$ are identical. Moreover, we observe that the two types of settings result in two distinct classes considering both performance measures, with the KP versions exhibiting a superior WNG performance to the non-KP versions but inferior DF performance. This separation is more dominant with $\mathbf{h}_{\mathrm{MWNG/MDF}}$ than with $\mathbf{h}_{\mathrm{MDF}}$, in particular in higher frequencies. In all cases, increasing M improves both performance measures.

Let us turn to the MFBR differential KP beamformer, $\mathbf{h}_{\mathrm{MFBR}}$. We maintain the same array settings used with $\mathbf{h}_{\mathrm{MDF}}$ and $\mathbf{h}_{\mathrm{MWNG/MDF}}$, and set a regularization factor of 10^{-2} with $\mathbf{\Gamma}_{\mathrm{b},\mathbf{a}}$ and $\mathbf{\Gamma}_{\mathrm{b},\mathbf{b}}$. As demonstrated in Fig. 6, the non-KP versions of $\mathbf{h}_{\mathrm{MFBR}}$ are of a better DF performance than its KP versions but of a worse WNG, particularly in high frequencies. The larger M, the lower the frequency above which the former is true. We deduce that with $\mathbf{h}_{\mathrm{MFBR}}$, in a similar manner to $\mathbf{h}_{\mathrm{MDF}}$ and $\mathbf{h}_{\mathrm{MWNG/MDF}}$, the flexibility in setting $M_{\mathbf{a}}$ and $M_{\mathbf{b}}$ allows one performance measure to improve at the expense of the other.

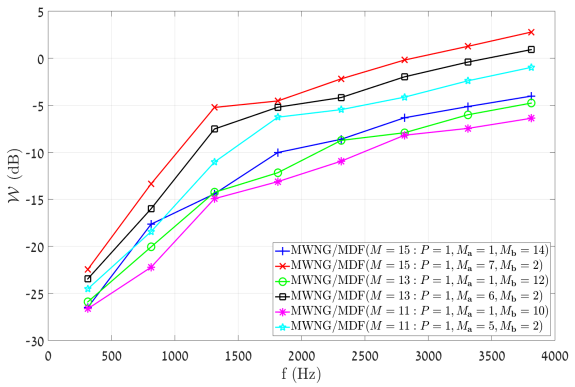


(a)

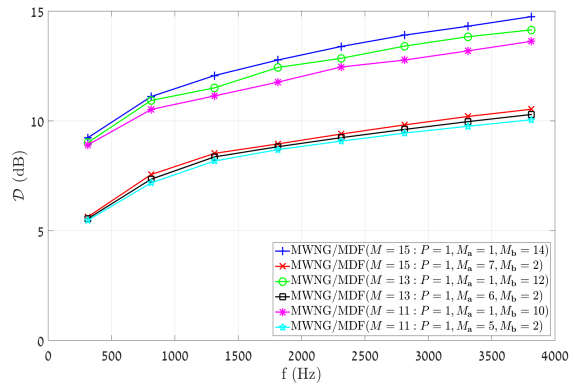


(b)

Fig. 4: WNG and DF measures with the MDF differential KP beamformer, $\mathbf{h}_{\mathrm{MDF}}$, with different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation parameters: $n = 4$, $\delta = 1$ cm and $\lambda = 10^{-2}$. (a) WNG and (b) DF.



(a)



(b)

Fig. 5: WNG and DF measures with the MWNG/MDF differential KP beamformer, $\mathbf{h}_{\mathrm{MWNG/MDF}}$, with different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$ and a varying number of microphones M. Simulation parameters: $\delta = 1$ cm and $\lambda = 10^{-3}$. (a) WNG and (b) DF.

B. Speech Signals Simulations in Reverberant Environments

In this part, we demonstrate the performances of differential KP beamformers on speech signals in practical simulated scenarios and reverberant environments. We investigate and compare the performances of all the beamformers presented in the paper with four distinct array settings of a $M = 9$ microphone array.

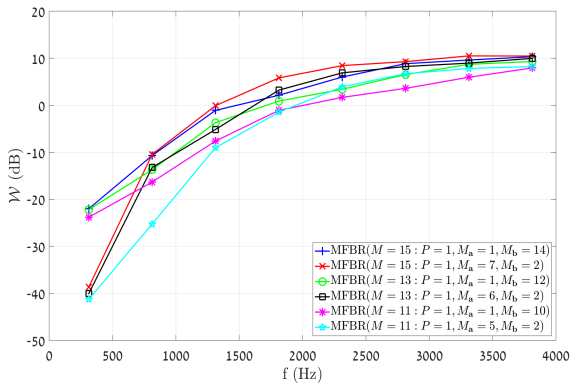
The reverberant simulations are performed as follows. We use a room impulse response (RIR) generator [35] to simulate the reverberant noise-free signal received in each of the microphones. The RIR generator is based on the image method of Allen and Berkley [36]. We simulate a $6 \times 6 \times 3$ m room in which a desired speech signal source is located at $(x, y, z) = (1, 1, 1.5)$ and an uncorrelated directional interference is located at $(x, y, z) = (3, 3, 1.5)$. 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 [37]. It is sampled at a sampling rate of $f_{\mathrm{s}} = 1/T_{\mathrm{s}} = 16$ kHz within the signal duration T. A ULA consisting of $M = 9$ microphones is located on the $(1, y, 1.5)$-axis, with $y = 2.96 : 3.04$. In addition to the directional interference, two uncorrelated noise fields are present: a white thermal Gaussian noise and a spherically isotropic diffuse noise. The latter and the directional interference are equally powerful, whereas the white thermal noise is 20 dB weaker than each of them. Denoting the combined noise signal at the reference (first) microphone in the time domain by $v(t)$, we may define the time-domain SNR (which is identical to the broadband SNR) by

$$\mathrm{iSNR}_{\mathrm{t}} = \frac{\int_t x^2(t)dt}{\int_t v^2(t)dt}, \tag{80}$$

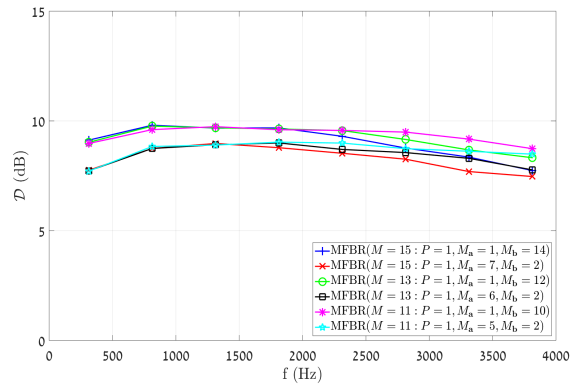
which is set to $\mathrm{iSNR}_{\mathrm{t}} = 0$ dB.

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). Next, differential KP beamformers with different array settings are independently applied to the noisy signal to yield clean signal estimates in the STFT domain, followed by an inverse STFT procedure to obtain time-domain enhanced signals. The latter is carried out by using the overlap-and-add method.

We simulate three reverberant scenarios with $T_{60} \in \{130, 250, 400\}$ msec, where T_{60} is defined by Sabine-Franklin's formula [38]. In each scenario, we design each of the five beamformers presented in the paper with five different



(a)



(b)

Fig. 6: WNG and DF measures with the MFBR differential KP beamformer, $\mathbf{h}_{\mathrm{MFBR}}$, with different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation parameters: $n = 4$ and $\delta = 1$ cm. (a) WNG and (b) DF.

array settings: $(0,1,9)$- a "classical" non-differential non-KP beamformer, $(0,3,3)$- a non-differential KP beamformer [27], $(1,1,8)$- a differential non-KP beamformer [24], and $(1,2,4)$ and $(1,4,2)$- two distinct differential KP beamformers. The NS differential KP beamformer $\mathbf{h}_{\mathrm{NS}}$ is designed with a single unique null in the direction of the directional interference.

Next, we are interested in objectively quantifying the performance of each of the five beamformers with the five aforementioned array settings. We shall do that by individually examining the power ratio of the following components of the time-domain enhanced signals and the noisy observation signals of the first microphone: the white thermal Gaussian noise, the diffuse noise, the reverberant directional interference and the desired speech signal reverberations. Formulating the noisy observation signal in the time-domain in microphone m, we have

$$\begin{aligned} y_m &= x_{\mathrm{d}} * g_{\mathrm{d},m} + v_{\mathrm{i}} * g_{\mathrm{i},m} + v_{\mathrm{d},m} + v_{\mathrm{w},m} \\ &= x_m + x_{\mathrm{r},m} + v_{\mathrm{r},m} + v_{\mathrm{d},m} + v_{\mathrm{w},m}, \end{aligned} \tag{81}$$

where $*$ is the linear convolution operator, x_{d} is the desired speech signal, v_{i} is the directional interference, $v_{\mathrm{d},m}$ and $v_{\mathrm{w},m}$ are, respectively, the additive diffuse noise and white noise in microphone m, $g_{\mathrm{d},m}$ is the RIR from the desired signal to microphone m, $g_{\mathrm{i},m}$ is the RIR from the directional interference to microphone m, and x_m, $x_{\mathrm{r},m}$, and $v_{\mathrm{r},m}$ are the direct path desired signal, its reverberations and the reverberant interference as received in microphone m, respectively. Additionally, we define the same components of the second row of (81) with respect to the time-domain enhanced speech signals by using the subscript f. For example, x_{f} is the enhanced direct path desired signal and $v_{\mathrm{w,f}}$ is the white noise component in the enhanced speech signal. We note that the energies of the direct path desired signal component in the first microphone and the enhanced direct path desired signal were verified to equal up to a scale of 0.1 dB in all scenarios and with all the beamformers.

We now address the white thermal noise and the diffuse noise. Using the notations of (81), we define the diffuse noise reduction (DNR) factor by

$$\mathrm{DNR} = \frac{E\left[v_{\mathrm{d},1}^2\right]}{E\left[v_{\mathrm{d,f}}^2\right]} \tag{82}$$

and the white noise reduction (WNR) factor by

$$\mathrm{WNR} = \frac{E\left[v_{\mathrm{w},1}^2\right]}{E\left[v_{\mathrm{w,f}}^2\right]}. \tag{83}$$

The DNR and WNR with each of the presented beamformers and the discussed array settings appear in Table I. To begin with, we note that, broadly, the $(0,1,9)$- and $(0,3,3)$-beamformers attain a better (higher) WNR but a worse (lower) DNR (except with $\mathbf{h}_{\mathrm{MDF}}$). Clearly, this is due to their non-differential nature. Focusing on the three other array settings, we observe that the KP beamformers exhibit a preferable WNR with $\mathbf{h}_{\mathrm{MDF}}$, $\mathbf{h}_{\mathrm{MWNG/MDF}}$, and $\mathbf{h}_{\mathrm{MFBR}}$, hence mitigating the white noise amplification issue of differential beamformers. This is with accordance to the WNG performance addressed in the previous sub-section and is considerably more significant with $(1,4,2)$. On the other hand, the DNR with $(1,1,8)$ and these three beamformers is equal to or better than with the KP settings. Turning to the $\mathbf{h}_{\mathrm{MWNG}}$ and $\mathbf{h}_{\mathrm{NS}}$, we observe that the KP beamformers exhibit a preferable DNR but worse WNR, with the performance gaps being more significant with the latter beamformer. We infer that the differential KP approach constitute a mean to flexibly tune the desirable performance trade-off of mitigating the white noise amplification and improving the diffuse noise attenuation.

Next, we define the desired signal reverberations reduction (RR) factor by

$$\mathrm{RR} = \frac{E\left[x_{\mathrm{r},1}^2\right]}{E\left[x_{\mathrm{r,f}}^2\right]} \tag{84}$$

and the interference reduction (IR) factor by

$$\mathrm{IR} = \frac{E\left[v_{\mathrm{r},1}^2\right]}{E\left[v_{\mathrm{r,f}}^2\right]}. \tag{85}$$

We note that both factors are a function of the RIRs, as opposed to the DNR and WNR. The RR and IR with the discussed beamformers and array settings for $T_{60} = 130$ msec

are shown in Table II. It is stressed that except for with $\mathbf{h}_{\mathrm{NS}}$, the differential beamformers attain a significantly higher IR than their non-differential counterparts. In addition, with $\mathbf{h}_{\mathrm{NS}}$, the latter measure is considerably higher with $(1,2,4)$ and $(1,4,2)$ than with $(1,1,8)$, whereas with the other beamformers the IRs are roughly equal. Focusing on the RR, the differential beamformers exhibit a preferable performance, with the $(1,4,2)$ settings attaining the highest value with $\mathbf{h}_{\mathrm{NS}}$ and the $(1,1,8)$ settings attaining the highest values with $\mathbf{h}_{\mathrm{MDF}}$, $\mathbf{h}_{\mathrm{MWNG/MDF}}$, and $\mathbf{h}_{\mathrm{MFBR}}$. The RR performance with $\mathbf{h}_{\mathrm{MWNG}}$ is equal with the three differential settings.

We turn to the $T_{60} = 250$ msec scenario for which the RR and IR are shown in Table III. Firstly, we observe that, in general, the RR increases and the IR decreases in comparison to the former scenario. This may be explained as follows. In a non-reverberant scenario, the RR is zero (which is low) and it is straightforward to attain a very high value of the IR (for example, by placing a null in the appropriate direction). As T_{60} increases, the desired signal reverberations are more paramount and, therefore, there is more room for a beamformer to reject these reverberations, implying a higher possible RR value. On the other hand, as T_{60} increases, the reverberations of the directional interference are greater scattered across the azimuth angle. Thus, the task of rejecting the spatial interference turns more complex. In spite of this observation, we note that the performance analysis of the previous scenario remains valid, with the values of the RR and IR change as explained but their corresponding differences with varying settings and fixed beamformers are maintained.

Lastly, we relate the $T_{60} = 400$ msec scenario of Table IV. We observe that indeed, according to the formerly discussed trend, the RR values are higher but the IR values are lower. In addition, the performance analysis of the two previous scenarios applies for this scenario as well.

We end this part by examining the same three reverberant scenarios discussed above from a different perspective. That is, we analyze the average PESQ [39] and STOI [40] scores of the time-domain enhanced speech signals. The results are depicted in Fig 7. It is clear that in terms of the PESQ score the $(1,4,2)$-beamformers are superior to their counterparts, excluding the $\mathbf{h}_{\mathrm{MWNG}}$. Particularly, the performance gap is the most significant in the $T_{60} = 130$ msec scenario. We observe that in the three scenarios $\mathbf{h}_{\mathrm{NS}}$ exhibits the highest overall PESQ score with the aforementioned settings, which is of a great contrast to the low PESQ scores of the same beamformer with $(1,1,8)$. Considering the reduction factors from the previous part, these results may be explained as follows. When T_{60} is low, the reverberations of the desired signal and interference are less significant. Hence, the white noise amplification becomes more considerable. As it is mitigated with the KP settings with $\mathbf{h}_{\mathrm{MDF}}$, $\mathbf{h}_{\mathrm{MWNG/MDF}}$, and $\mathbf{h}_{\mathrm{MFBR}}$ with respect to the non-KP settings, their corresponding time-domain enhanced signals are of a higher quality. On the contrary, while with the $(1,4,2)$ version of $\mathbf{h}_{\mathrm{NS}}$ the white noise is even greater amplified, its preferable DNR, RR, and IR values with respect to most or all of its counterparts result in higher quality enhanced signals. Focusing on the STOI scores in the $T_{60} = 130$ msec scenario, we observe that indeed, the beamformers with the differential KP settings, that is, $(1,2,4)$ and $(1,4,2)$, are of a better intelligibility, with the latter outperforming the former. Increasing T_{60} to 250 msec reduces the performance gap, whereas with $T_{60} = 400$ msec the STOI scores are roughly equal with most combinations of beamformers and array settings.

Considering all the presented beamformers and array settings, it is beneficial to conclude by proposing some design rules of thumbs. For starters, one must choose an appropriate beamformer type. For example, in case powerful directional interferences are present $\mathbf{h}_{\mathrm{NS}}$ is likely to be preferred, and in case it is desirable to substantially attenuate the array response in the rear-half plane $\mathbf{h}_{\mathrm{MFBR}}$ should be chosen. Alternatively, if the top-priority attribute of the array is white-noise attenuation $\mathbf{h}_{\mathrm{MWNG}}$ is an appropriate choice, whereas $\mathbf{h}_{\mathrm{MWNG/MDF}}$ and $\mathbf{h}_{\mathrm{MDF}}$ should be considered if the array directivity is of the highest significance. Then, one should determine the array settings, with P being the more dominant parameter, whereas $M_{\mathbf{a}}$ and $M_{\mathbf{b}}$ enable a finer tuning. That is, the value of P sets a level of array directivity at the expanse of white-noise sensitivity. Depending on the selection of beamformer type, the values of $M_{\mathbf{a}}$ and $M_{\mathbf{b}}$ may either improve the array directivity or white-noise attenuation. Nevertheless, the high flexibility of this approach requires a careful design, as it is not guaranteed that every differential KP combination of the array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$, that is, $P > 0$ and $M_{\mathbf{a}}, M_{\mathbf{b}} > 1$, would yield a better choice than non-differential ($P = 0$) or non-KP ($M_{\mathbf{a}} = 1$ or $M_{\mathbf{b}} = 1$) settings.

C. Reverberant Simulations with Array Imperfections

In this part, we focus on the effects of two common types of array imperfections: deviation of the speech source incident angle and misplacements of the array microphones. We maintain the same simulation settings of Section VI-B, set $T_{60} = 130$ msec, and analyze the performance of a subset of the previously discussed beamformers with two array sizes.

Let us describe the two types of array imperfections. To begin with, we move the desired speech signal source along the x-axis to generate an incident angle of $\theta_{\mathrm{s}} = 10^{o}$. In addition, we simulate microphones misplacements by adding independent normally-distributed values of zero mean and standard deviation of 1 mm to each of the microphones positions along the y-axis. Then, we compare the three following scenarios.

- Scen. (a): no array imperfections (reference scenario).
- Scen. (b): microphones misplacements but true speech source incident angle.
- Scen. (c): microphones misplacements and deviation of the speech source incident angle.

We examine the PESQ and STOI scores in each of these scenarios with $\mathbf{h}_{\mathrm{MWNG}}$, $\mathbf{h}_{\mathrm{MDF}}$ and $\mathbf{h}_{\mathrm{MWNG/MDF}}$ with two array sizes: $M = 5$ and $M = 9$. In each, we show examples of non-differential non-KP beamformers, differential non-KP beamformers and differential KP beamformers. We note that with the $M = 9$ array we maintain the same array settings as above and show a subset of three out of five of them: $(0,1,9)$, $(1,1,8)$, and $(1,4,2)$. Additionally, we employ a

TABLE I: The DNR and WNR for Differential KP Beamformers with Different Array Settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation Parameters: $M = 9$, $\delta = 1$ cm, $n = 4$, and $\mathrm{iSNR}_{\mathrm{t}} = 0$ dB.

	DNR (dB)					WNR (dB)				
Settings	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$
$\mathbf{h}_{\mathrm{MWNG}}$	4.6	4.6	6.3	7.0	6.8	**9.5**	**9.5**	$\mathbf{-1.4}$	$\mathbf{-3.5}$	$\mathbf{-3.0}$
$\mathbf{h}_{\mathrm{NS}}$	6.3	7.3	7.5	8.5	10.1	-1.4	-11.0	-22.9	-26.0	-28.6
$\mathbf{h}_{\mathrm{MDF}}$	**10.7**	**10.4**	**11.8**	**11.1**	**10.4**	-2.4	-4.6	-23.4	-21.7	-15.4
$\mathbf{h}_{\mathrm{MWNG/MDF}}$	**10.7**	7.9	**11.8**	10.4	8.9	-2.4	-0.9	-23.4	-21.4	-15.6
$\mathbf{h}_{\mathrm{MFBR}}$	5.5	4.7	8.0	8.0	7.4	2.8	2.3	-24.7	-23.8	-17.9

TABLE II: The RR and IR for Differential KP Beamformers with Different Array Settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation Parameters: $T_{60} = 130$ msec, $M = 9$, $\delta = 1$ cm, $n = 4$, and $\mathrm{iSNR}_{\mathrm{t}} = 0$ dB.

	RR (dB)					IR (dB)				
Settings	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$
$\mathbf{h}_{\mathrm{MWNG}}$	0.2	0.2	3.2	3.3	3.3	6.2	6.2	**20.7**	19.7	20.4
$\mathbf{h}_{\mathrm{NS}}$	3.2	3.6	5.5	3.9	**6.4**	**20.7**	**21.1**	17.5	20.5	20.0
$\mathbf{h}_{\mathrm{MDF}}$	**4.0**	**4.4**	**5.8**	**5.3**	4.4	16.1	15.1	**20.7**	**20.9**	**21.0**
$\mathbf{h}_{\mathrm{MWNG/MDF}}$	**4.0**	2.4	**5.8**	**5.3**	4.2	16.1	12.6	**20.7**	20.6	20.5
$\mathbf{h}_{\mathrm{MFBR}}$	2.6	2.3	5.0	4.8	4.2	14.3	13.0	19.6	19.8	19.9

TABLE III: The RR and IR for Differential KP Beamformers with Different Array Settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation Parameters: $T_{60} = 250$ msec, $M = 9$, $\delta = 1$ cm, $n = 4$, and $\mathrm{iSNR}_{\mathrm{t}} = 0$ dB.

	RR (dB)					IR (dB)				
Settings	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$
$\mathbf{h}_{\mathrm{MWNG}}$	0.5	0.5	4.2	4.4	4.3	3.5	3.5	9.7	8.6	9.6
$\mathbf{h}_{\mathrm{NS}}$	4.2	4.7	6.1	**6.8**	**7.9**	**9.7**	**10.2**	6.8	10.6	9.9
$\mathbf{h}_{\mathrm{MDF}}$	**5.4**	**5.7**	**7.1**	6.7	5.7	9.3	9.0	**10.9**	**11.0**	**11.2**
$\mathbf{h}_{\mathrm{MWNG/MDF}}$	**5.4**	3.5	**7.1**	6.6	5.4	9.3	7.0	**10.9**	10.7	10.2
$\mathbf{h}_{\mathrm{MFBR}}$	3.8	3.5	6.3	6.2	5.4	6.3	5.6	8.7	9.0	9.0

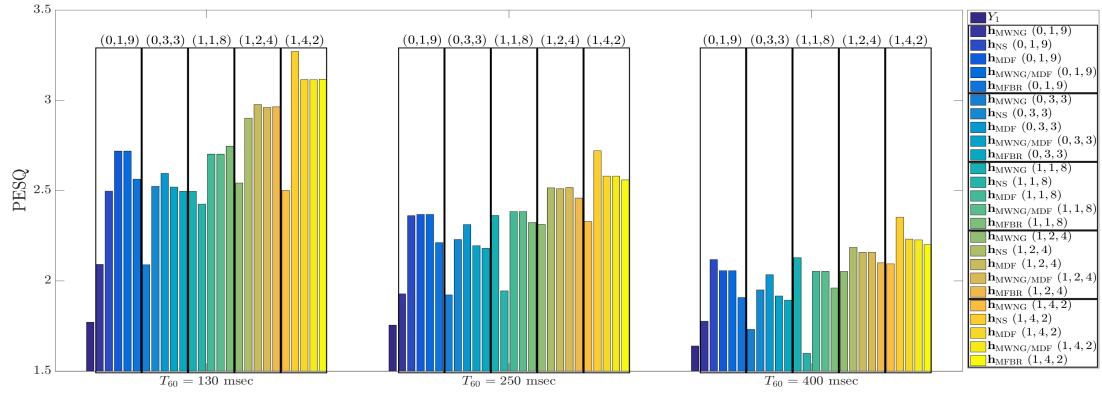
TABLE IV: The RR and IR for Differential KP Beamformers with Different Array Settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation Parameters: $T_{60} = 400$ msec, $M = 9$, $\delta = 1$ cm, $n = 4$, and $\mathrm{iSNR}_{\mathrm{t}} = 0$ dB.

	RR (dB)					IR (dB)				
Settings	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$	$(\mathbf{0,1,9})$	$(\mathbf{0,3,3})$	$(\mathbf{1,1,8})$	$(\mathbf{1,2,4})$	$(\mathbf{1,4,2})$
$\mathbf{h}_{\mathrm{MWNG}}$	0.6	0.6	5.1	5.4	5.3	1.8	1.8	6.8	5.8	6.4
$\mathbf{h}_{\mathrm{NS}}$	5.1	5.7	6.6	7.7	**8.8**	**6.6**	**7.3**	4.4	7.5	6.8
$\mathbf{h}_{\mathrm{MDF}}$	**6.6**	**6.9**	**8.3**	**7.8**	6.8	6.2	6.0	**8.0**	**7.9**	**7.7**
$\mathbf{h}_{\mathrm{MWNG/MDF}}$	**6.6**	4.4	**8.3**	**7.8**	6.5	6.2	4.5	**8.0**	7.7	7.0
$\mathbf{h}_{\mathrm{MFBR}}$	4.9	4.6	7.6	7.4	6.5	3.5	2.9	5.9	6.2	5.8

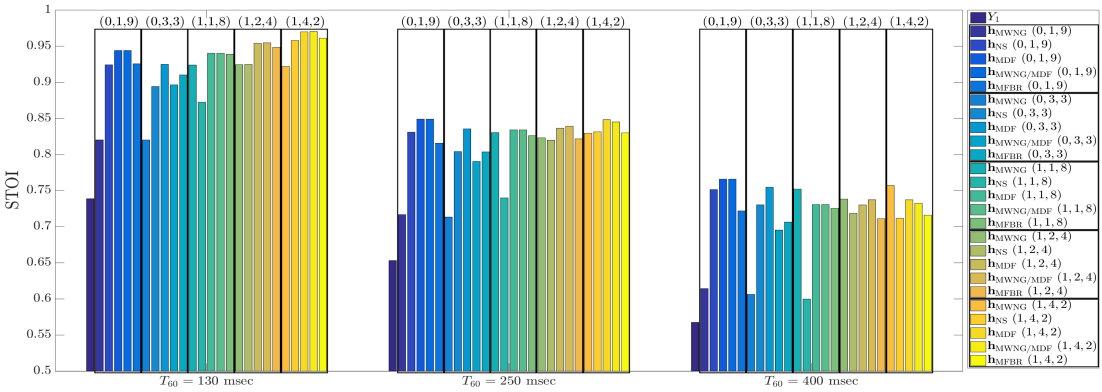
$M = 5$ array (formed by dropping the the first-two and last-two microphones of the $M = 9$ array) as a practical example for small-size arrays. With the latter, we use the following array settings: $(0, 1, 5)$, $(1, 1, 4)$, and $(1, 2, 2)$.

The PESQ and STOI scores with the $M = 5$ and $M = 9$ arrays are shown, respectively, in Fig. 8 and Fig. 9. To begin with, we observe that in the reference scenario, i.e., Scen. (a), and with both arrays, the beamformers with the differential KP settings outperform their counterparts in terms of both the PESQ and STOI scores. In particular, the performance gap is accentuated with $\mathbf{h}_{\mathrm{MDF}}$ and $\mathbf{h}_{\mathrm{MWNG/MDF}}$. As we consider the array imperfections of Scen. (b) and Scen. (c), we note that while the performances of all the presented beamformers slightly degrade- the performance gap remains. We infer that even with small arrays and practical imperfections the differential KP approach may outperform the rest of the discussed approaches.

Finally, for the sake of completeness, we address the beampattern measure. As an example, we focus on the MWNG differential KP beamformer, $\mathbf{h}_{\mathrm{MWNG}}$, with the six combinations of array settings and sizes discussed in this part. The beampatterns are plotted in Fig. 10. We observe that with the $M = 5$ array the $(0, 1, 5)$ version of $\mathbf{h}_{\mathrm{MWNG}}$ exhibits a supercardioid-like shape, whereas its $(1, 1, 4)$ and $(1, 2, 2)$ versions exhibit a dipole-like shape. Nevertheless, we observe that the back lobe with the latter is roughly 5 dB lower than with the former, indicating a preferable directivity. Examining the $M = 9$ array, we note that the beampatterns of the $(0, 1, 9)$ and $(1, 1, 8)$ versions are, roughly, similarly-shaped, but the latter offers a preferable directivity. On the contrary, the



(a)



(b)

Fig. 7: Average PESQ and STOI scores of enhanced speech signals with the five differential KP beamformers presented in the paper and with five different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation parameters: $M = 9$, $\delta = 1$ cm, and $\mathrm{iSNR}_{\mathrm{t}} = 0$ dB. (a) PESQ scores and (b) STOI scores.

$(1, 4, 2)$ version exhibits a different beampattern shape whose side lobe is significantly lower than the side lobes of the over versions- a paramount attribute in scenarios of substantial early reverberations.

VII. Conclusions

We have generalized the multistage differential beamforming approach by applying a KP decomposition to a global differential beamformer, and independently optimizing the two sub-beamformers. Previous non-differential or non-KP beamformers may be obtained by an appropriate selection of the array settings parameters. We proposed five types of differential KP beamformers and demonstrated that each may perform better than previous approaches in terms of the WNG or DF measure at the expense of the complementary measure, depending on the array settings. This flexibility enables one to mitigate the white noise amplification with the differential MDF, MWNG/MDF, and MFBR beamformers or improve the directivity with the differential MWNG and NS beamformers. In addition, we showed that signal reverberations are attenuated to the greatest extent using the NS differential KP beamformer, whereas reverberations of a directional interference are equally attenuated using the other beamformers with differential KP and differential non-KP settings. Finally, we examined the average PESQ and STOI scores of the respective time-domain enhanced signals and demonstrated that both are higher with the new approach, even under array imperfections. This is in particular true for moderately reverberant environments.

Acknowledgement

The authors thank the anonymous reviewers for their constructive comments and helpful suggestions.

References

[1] G. W. Elko, "Microphone array systems for hands-free telecommunication," *Speech Communication*, vol. 20, no. 3, pp. 229 – 240, 1996, Acoustic Echo Control and Speech Enhancement Techniques.

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

[3] 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.

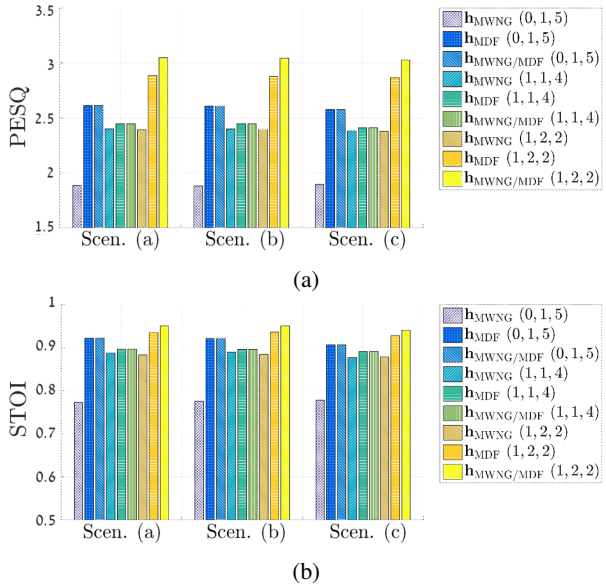


Fig. 8: Average PESQ and STOI scores of enhanced speech signals in the three array imperfection scenarios described in the paper. We simulate three differential KP beamformers with three different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation parameters: $M = 5$, $T_{60} = 130$ msec, $\delta = 1$ cm, and $\text{iSNR}_{\text{t}} = 0$ dB. (a) PESQ scores and (b) STOI scores.

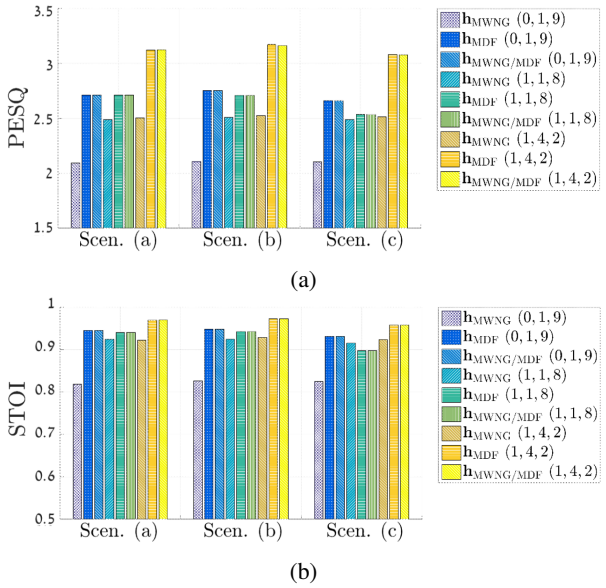


Fig. 9: Average PESQ and STOI scores of enhanced speech signals in the three array imperfection scenarios described in the paper. We simulate three differential KP beamformers with three different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$. Simulation parameters: $M = 9$, $T_{60} = 130$ msec, $\delta = 1$ cm, and $\text{iSNR}_{\text{t}} = 0$ dB. (a) PESQ scores and (b) STOI scores.

[4] G. Huang, J. Chen, and J. Benesty, "On the design of differential beamformers with arbitrary planar microphone array geometry," *The Journal of the Acoustical Society of America*, vol. 144, no. 1, pp. EL66–EL70, 2018.

[5] J. Weinberger, H.F. Olson, and F. Massa, "A uni-directional ribbon microphone," *The Journal of the Acoustical Society of America*, vol. 5, no. 2, pp. 139–147, 1933.

[6] H. F. Olson, "Gradient microphones," *The Journal of the Acoustical Society of America*, vol. 17, no. 3, pp. 192–198, 1946.

[7] G. W. Elko, *Differential Microphone Arrays*, pp. 11–65, Springer US, Boston, MA, 2004.

[8] M. Kolundzija, C. Faller, and M. Vetterli, "Spatiotemporal gradient analysis of differential microphone arrays," *journal of the audio engineering society*, vol. 59, no. 1, pp. 20–28, Jan. 2011.

[9] Y. Buchris, I. Cohen, and J. Benesty, "On the design of time-domain differential microphone arrays," *Applied Acoustics*, vol. 148, pp. 212 – 222, 2019.

[10] J. Benesty, I. Cohen, and J. Chen, *Array Beamforming with Linear Difference Equations*, Springer, 2021.

[11] J. Jin, G. Huang, X. Wang, J. Chen, J. Benesty, and I. Cohen, "Steering study of linear differential microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 29, pp. 158–170, 2021.

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

[13] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*, Springer-Verlag Berlin Heidelberg, 2013.

[14] T. Long, J. Benesty, J. Chen, and I. Cohen, "Differential beamformers derived from approximate performance measures," in *2018 16th International Workshop on Acoustic Signal Enhancement (IWAENC)*, 2018, pp. 66–70.

[15] A. Bernardini, F. Antonacci, and A. Sarti, "Wave digital implementation of robust first-order differential microphone arrays," *IEEE Signal Processing Letters*, vol. 25, no. 2, pp. 253–257, 2018.

[16] F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, "Uniform linear arrays of first-order steerable differential microphones," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 12, pp. 1906–1918, 2019.

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

[18] E. De Sena, H. Hacihabiboglu, and Z. Cvetkovic, "On the Design and Implementation of Higher Order Differential Microphones," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 1, pp. 162–174, 2012.

[19] M. Buck, "Aspects of first-order differential microphone arrays in the presence of sensor imperfections," *European Transactions on Telecommunications*, vol. 13, pp. 115–122, 2002.

[20] X. Wu, H. Chen, J. Zhou, and T. Guo, "Study of the mainlobe misorientation of the first-order steerable differential array in the presence of microphone gain and phase errors," *IEEE Signal Processing Letters*, vol. 21, no. 6, pp. 667–671, 2014.

[21] X. Wu and H. Chen, "Directivity Factors of the First-Order Steerable Differential Array With Microphone Mismatches: Deterministic and Worst-Case Analysis," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 24, no. 2, pp. 300–315, 2016.

[22] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*, Springer-Verlag Berlin Heidelberg, Berlin, 1st edition, 2009.

[23] S. He and H. Chen, "Closed-form DOA estimation using first-order differential microphone arrays via joint temporal-spectral-spatial processing," *IEEE Sensors Journal*, vol. 17, no. 4, pp. 1046–1060, 2017.

[24] G. Huang, J. Benesty, I. Cohen, and J. Chen, "A simple theory and new method of differential beamforming with uniform linear microphone arrays," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1079–1093, 2020.

[25] Y. I. Abramovich, G. J. Frazer, and B. A. Johnson, "Iterative Adaptive Kronecker MIMO Radar Beamformer: Description and Convergence Analysis," *IEEE Transactions on Signal Processing*, vol. 58, no. 7, pp. 3681–3691, 2010.

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

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

[28] I. Cohen, J. Benesty, and J. Chen, "Differential kronecker product beamforming," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 27, no. 5, pp. 892–902, 2019.

[29] G. Huang, J. Benesty, J. Chen, and I. Cohen, "Robust and steerable kronecker product differential beamforming with rectangular microphone arrays," in *ICASSP 2020 - 2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, 2020, pp. 211–215.

[30] G. Huang, I. Cohen, J. Benesty, and J. Chen, "Kronecker product beamforming with multiple differential microphone arrays," in *2020*

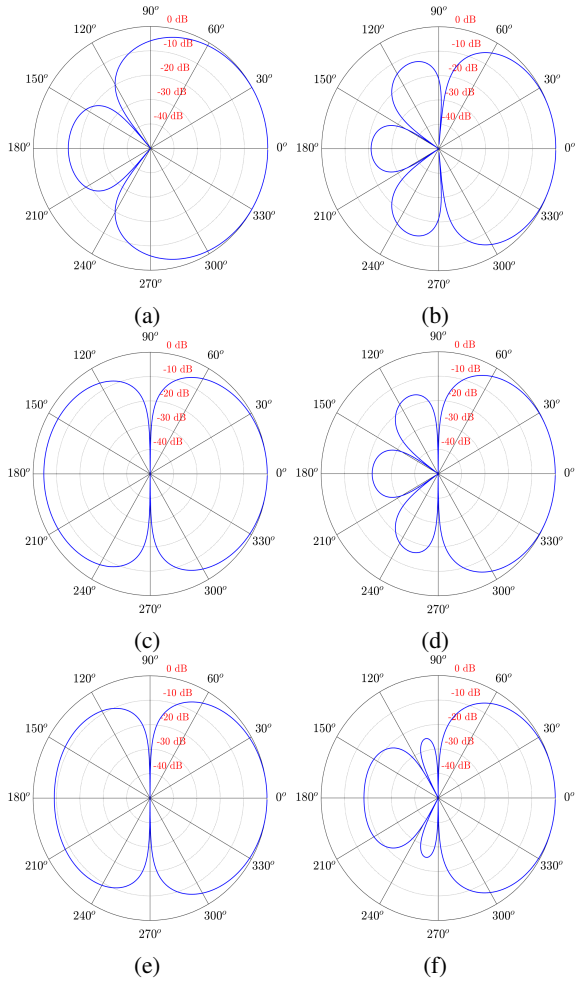


Fig. 10: Beampatterns of the MWNG differential KP beamformer, $\mathbf{h}_{\mathrm{MWNG}}$, with six different array settings $(P, M_{\mathbf{a}}, M_{\mathbf{b}})$, and two values of M. Simulation parameters: $f = 2$ kHz and $\delta = 1$ cm. (a) $M = 5$; $(P = 0, M_{\mathbf{a}} = 1, M_{\mathbf{b}} = 5)$, (b) $M = 9$; $(P = 0, M_{\mathbf{a}} = 1, M_{\mathbf{b}} = 9)$, (c) $M = 5$; $(P = 1, M_{\mathbf{a}} = 1, M_{\mathbf{b}} = 4)$, (d) $M = 9$; $(P = 1, M_{\mathbf{a}} = 1, M_{\mathbf{b}} = 8)$; (e) $M = 5$; $(P = 1, M_{\mathbf{a}} = 2, M_{\mathbf{b}} = 2)$, and (f) $M = 9$; $(P = 1, M_{\mathbf{a}} = 4, M_{\mathbf{b}} = 2)$.

IEEE 11th Sensor Array and Multichannel Signal Processing Workshop (SAM), 2020, pp. 1–5.

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

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

[33] F. Yates, "The Analysis of Replicated Experimetns when the Field Results are Incomplete," *Empire Journal of Experimental Agriculture*, vol. 1, no. 2, pp. 129–142, 1933.

[34] H. Wold and F. Lyttkens, "Nonlinear iterative partial least squares (NIPALS) estimation procedures," *Bulletin of the International Statistical Institute*, vol. 43, no. 1, pp. 29–47, 1969.

[35] E. A. P. Habets, "Rir-generator," https://www.audiolabs-erlangen.de/fau/professor/habets/software/rir-generator.

[36] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, 1979.

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

[38] A. Pierce, *Acoustics: An Introduction to Its Physical Principles and Applications*, Acoustical Society of America, 1991.

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

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