A Simple Theory and New Method of Differential Beamforming with Uniform Linear Microphone Arrays

Gongping Huang, Jacob Benesty, Israel Cohen, and Jingdong Chen

***Abstract*—This paper presents a theoretical study of differential beamforming with uniform linear arrays. By defining a forward spatial difference operator, any order of the spatial difference of the observed signals can be represented as a product of a difference operator matrix and the microphone array observations. Consequently, differential beamforming is implemented in two stages, where the first one obtains spatial difference of the observations and the second stage optimizes the beamformer. The major contributions of this paper are as follows. First, we propose a new theory of differential beamforming with uniform linear arrays, which shows clearly the connection between the conventional differential beamforming and the null-constrained differential beamforming methods. This provides some new insight into the design of differential beamformers. Second, we deduce some new differential beamformers, where conventional beamforming may be seen as a particular case. Specifically, we derive the maximum white noise gain (MWNG), maximum directivity factor (MDF), parameterized MDF, and parameterized maximum front-to-back ratio differential beamformers. Third, we further extend the idea of how to design optimal differential beamformers by combining both the observed signals and their spatial differences.**

***Index Terms*—Microphone arrays, uniform linear arrays, differential beamforming, fixed beamformer, forward spatial difference operator.**

I. Introduction

Microphone arrays with proper beamforming techniques are widely used in voice communication and human-machine speech interface systems to enhance a signal of interest from its noisy observations [1]–[6]. Among numerous beamforming methods developed in the literature, differential microphone arrays have attracted a significant amount of interest [7]–[12]. Generally, differential microphone arrays refer to arrays that combine closely spaced sensors to respond to the spatial derivatives of the acoustic pressure field [13]–[18], so they are usually small in size and easy to be integrated into small communication devices such as wearable devices and intelligent speakers [19]–[21]. Conventionally, differential microphone arrays are designed in a multistage manner, which measure the differentials of the acoustic pressure field by combining the outputs of a number of omnidirectional sensors [14], [22]. Specifically, the output of a first-order differential beamformer is the difference between the outputs of two adjacent omnidirectional microphones, and the output of an nth-order ($n \geq 1$) differential beamformer is the difference between the outputs of two $(n-1)$th-order differential beamformers [10], [22], [23].

However, conventional differential beamformers are very sensitive to sensor noise and array imperfections at low frequencies, which restricts to a certain degree their practical application [13], [16], [24]. To overcome this drawback, a method was developed to form differential beamformers in the short-time Fourier transform (STFT) domain, by solving a linear system constructed from null constraints [16], [25]. In this approach, the robustness of the beamformers, which is quantified by the so-called white noise gain (WNG), can be improved with the minimum-norm solution by increasing the number of microphones [16]. Alternatively, differential beamformers can be designed by finding the relations between the beamformer's beampattern and the desired directivity pattern with some series expansion approximations, such as Maclaurin series and Jacobi-Anger expansion series [26], [27] in the STFT domain. While they are flexible in forming different directivity patterns and improving the robustness of differential microphone arrays, these methods convert the differential beamforming problem into one of linear system solving or optimization, and the principle of differential operation becomes less obvious. Furthermore, the white noise amplification problem still exists at low frequencies even though it is greatly mitigated by increasing the number of microphones while fixing the differential microphone array order. As a result, further studies on differential microphone arrays are still indispensable.

In this paper, we present a new theory of differential beamforming with uniform linear arrays. We first define a forward spatial difference operator of the observed signals, where any order of the spatial difference of the observation signals can be represented as a product of a difference operator matrix and the microphone array observations. Differential beamforming is constructed in two main stages, i.e., differential stage and beamforming stage. In the differential stage, a Pth-order differential beamformer with M omnidirectional microphones yields $M-P$ spatially differential signals. In the beamforming stage, a linear filter of length $M-P$ is designed and applied to the differential signals to obtain the optimal beamforming performance. We show how to deduce new differential beamformers, where conventional beamforming can be viewed as a particular case. We derive the maximum WNG (MWNG) and maximum directivity factor (MDF) differential beamformers. Generally, the MDF beamformer will lead to a high value

This research was supported by the Israel Science Foundation (grant no. 576/16) and the ISF-NSFC joint research program (grant No. 2514/17 and 61761146001).

G. Huang and Israel Cohen are with Andrew and Erna Viterby Faculty of Electrical Engineering, Technion – Israel Institute of Technology, Technion City, Haifa 3200003, Israel (e-mail: gongpinghuang@gmail.com, 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: benesty@emt.inrs.ca).

J. Chen is with the Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an, Shaanxi 710072, China (e-mail: jingdongchen@ieee.org).

of the directivity factor (DF), but at the expense of white noise amplification, while the MWNG beamformer yields high WNG, but at the expense of low DF. To compromise between WNG and DF, we propose a parameterized MDF beamformer, where DF and WNG can be compromised by changing a parameter. In a similar way, we propose a parameterized maximum front-to-back ratio beamformer that can compromise among WNG, DF, and front-to-back ratio. Notice that some methods were developed in the literature to apply subspaces spanned by generalized eigenvectors of the signal/noise covariance matrices to adaptive beamforming, source separation, and speech enhancement to improve performance [28]–[31], which is, however, very much different of what is presented in this work.

The rest of the paper is organized as follows. In Section II, we present the signal model, the conventional beamforming, and some performance measures. In Section III, we present a simple theory of differential beamforming with ULAs. In Section IV, we develop a class of optimal differential beamformers. Finally, we present our simulation results in Section V, followed by some conclusions in Section VI.

II. Signal Model, Conventional Beamforming, and Performance Measures

We consider a source signal of interest (plane wave), in the farfield, that propagates from the azimuth angle, θ, in an anechoic acoustic environment at the speed of sound, i.e., $c = 340$ m/s, and impinges on a uniform linear array consisting of M (with $M \geq 2$) omnidirectional microphones. In this scenario, the corresponding steering vector (of length M) is [32]

$$\begin{aligned}\mathbf{d}_\theta(\omega) &= \left[\begin{array}{cccc} D_{\theta,1}(\omega) & D_{\theta,2}(\omega) & \cdots & D_{\theta,M}(\omega) \end{array}\right]^T \\ &= \left[\begin{array}{cccc} 1 & e^{-\jmath\varpi_\theta(\omega)} & \cdots & e^{-\jmath(M-1)\varpi_\theta(\omega)} \end{array}\right]^T, \end{aligned} \quad (1)$$

where $\jmath = \sqrt{-1}$ is the imaginary unit,

$$\varpi_\theta(\omega) = \frac{\omega\delta\cos\theta}{c}, \quad (2)$$

$\omega = 2\pi f$ is the angular frequency, $f > 0$ is the temporal frequency, δ is the interelement spacing, and the superscript ^{T} is the transpose operator.

Using the steering vector defined in (1), we can express the frequency-domain observed signal vector of length M as [34]

$$\begin{aligned}\mathbf{y}(\omega) &= \left[\begin{array}{cccc} Y_1(\omega) & Y_2(\omega) & \cdots & Y_M(\omega) \end{array}\right]^T \\ &= \mathbf{x}(\omega) + \mathbf{v}(\omega) \\ &= \mathbf{d}_\theta(\omega) X(\omega) + \mathbf{v}(\omega), \end{aligned} \quad (3)$$

where $Y_m(\omega)$ is the mth microphone signal, $\mathbf{x}(\omega) = \mathbf{d}_\theta(\omega) X(\omega)$, $X(\omega)$ is the zero-mean desired source signal, and $\mathbf{v}(\omega)$ is the zero-mean additive noise signal vector defined similarly to $\mathbf{y}(\omega)$. In the rest, in order to simplify the notation, we drop the dependence on the angular frequency, ω. From (3), it is clear that the $M \times M$ covariance matrix of $\mathbf{y}$ is

$$\mathbf{\Phi}_\mathbf{y} = E\left(\mathbf{y}\mathbf{y}^H\right) = \phi_X \mathbf{d}_\theta \mathbf{d}_\theta^H + \mathbf{\Phi}_\mathbf{v}, \quad (4)$$

where $E(\cdot)$ denotes mathematical expectation, the superscript ^{H} is the conjugate-transpose operator, $\phi_X = E\left(|X|^2\right)$ 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 the same at all sensors, i.e., $\phi_V = \phi_{V_1} = \phi_{V_2} = \cdots = \phi_{V_M}$, with $\phi_{V_m} = E\left(|V_m|^2\right)$, $m = 1, 2, \ldots, M$, we can express (4) as

$$\mathbf{\Phi}_\mathbf{y} = \phi_X \mathbf{d}_\theta \mathbf{d}_\theta^H + \phi_V \mathbf{\Gamma}_\mathbf{v}, \quad (5)$$

where $\mathbf{\Gamma}_\mathbf{v} = \mathbf{\Phi}_\mathbf{v}/\phi_V$ is the pseudo-coherence matrix of the noise. In the case of the spherically isotropic (diffuse) noise field, which will be very often assumed in this work, (5) becomes

$$\mathbf{\Phi}_\mathbf{y} = \phi_X \mathbf{d}_\theta \mathbf{d}_\theta^H + \phi_\mathrm{d} \mathbf{\Gamma}_\mathrm{d}, \quad (6)$$

where ϕ_d is the variance of the diffuse noise and $\mathbf{\Gamma}_\mathrm{d}$ is the pseudo-coherence matrix of the diffuse noise, whose (i,j)th $(i, j = 1, 2, \ldots, M)$ element is [33]

$$\left(\mathbf{\Gamma}_\mathrm{d}\right)_{ij} = \mathrm{sinc}\left[\varpi_0(i-j)\right], \quad (7)$$

with $\mathrm{sinc}(x) = \sin x/x$ and $\varpi_0 = \omega\delta/c$.

In order to be in the optimal working conditions of differential beamforming, we make the following two assumptions [16], [35].

(i) The sensor spacing, δ, is much smaller than the acoustic wavelength, $\lambda = c/f$, i.e., $\delta \ll \lambda$ (this implies that $\omega\tau_0 \ll 2\pi$). This assumption is required so that the true acoustic pressure differentials can be approximated by finite differences of the microphones' outputs.

(ii) The source signal of interest (also called the desired source signal) propagates from the angle $\theta = 0$ (endfire direction). Therefore, (3) becomes

$$\mathbf{y} = \mathbf{d}_0 X + \mathbf{v}, \quad (8)$$

and, at the endfire, the value of the beamformer beampattern should always be equal to 1 (or maximal).

Conventional beamforming is performed by applying a complex-valued linear filter, $\mathbf{h}$ of length M, to the observed signal vector, i.e.,

$$Z = \mathbf{h}^H\mathbf{y} = X_\mathrm{fd} + V_\mathrm{rn}, \quad (9)$$

where Z is the beamformer output signal (i.e., the estimate of X), $X_\mathrm{fd} = X\mathbf{h}^H\mathbf{d}_0$ is the filtered desired signal, and $V_\mathrm{rn} = \mathbf{h}^H\mathbf{v}$ is the residual noise. Since the two terms on the right-hand side of (9) are incoherent, the variance of Z is the sum of two variances:

$$\phi_Z = \mathbf{h}^H\mathbf{\Phi}_\mathbf{y}\mathbf{h} = \phi_{X_\mathrm{fd}} + \phi_{V_\mathrm{rn}}, \quad (10)$$

where $\phi_{X_\mathrm{fd}} = \phi_X\left|\mathbf{h}^H\mathbf{d}_0\right|^2$ and $\phi_{V_\mathrm{rn}} = \mathbf{h}^H\mathbf{\Phi}_\mathbf{v}\mathbf{h}$. In our context, the distortionless constraint is desired, i.e.,

$$\mathbf{h}^H\mathbf{d}_0 = 1. \quad (11)$$

Next, we give several important performance measures, which are always used to evaluate all kinds of conventional fixed and differential beamformers.

From (5), with $\theta = 0$, we can define the input signal-to-noise ratio (SNR) or the SNR corresponding to $\mathbf{y}$ as

$$\text{iSNR} = \text{SNR}_{\mathbf{y}} = \frac{\text{tr}\left(\phi_X \mathbf{d}_0 \mathbf{d}_0^H\right)}{\text{tr}\left(\phi_V \mathbf{\Gamma}_{\mathbf{v}}\right)} = \frac{\phi_X}{\phi_V}, \tag{12}$$

where $\text{tr}(\cdot)$ denotes the trace of a square matrix. According to (10), the output SNR can be defined as

$$\text{oSNR}\left(\mathbf{h}\right) = \phi_X \frac{\left|\mathbf{h}^H \mathbf{d}_0\right|^2}{\mathbf{h}^H \mathbf{\Phi}_{\mathbf{v}} \mathbf{h}} = \frac{\phi_X}{\phi_V} \times \frac{\left|\mathbf{h}^H \mathbf{d}_0\right|^2}{\mathbf{h}^H \mathbf{\Gamma}_{\mathbf{v}} \mathbf{h}}. \tag{13}$$

From the previous definitions of the SNRs, we deduce the array gain:

$$\mathcal{G}\left(\mathbf{h}\right) = \frac{\text{oSNR}\left(\mathbf{h}\right)}{\text{iSNR}} = \frac{\left|\mathbf{h}^H \mathbf{d}_0\right|^2}{\mathbf{h}^H \mathbf{\Gamma}_{\mathbf{v}} \mathbf{h}}. \tag{14}$$

A common way to evaluate the sensitivity of the array to some of its imperfections, such as sensors' self noise (which is white noise in nature), with a specific beamformer, is via the so-called white noise gain (WNG), which is defined by taking $\mathbf{\Gamma}_{\mathbf{v}} = \mathbf{I}_M$ in (14), where $\mathbf{I}_M$ is the $M \times M$ identity matrix, i.e.,

$$\mathcal{W}\left(\mathbf{h}\right) = \frac{\left|\mathbf{h}^H \mathbf{d}_0\right|^2}{\mathbf{h}^H \mathbf{h}}. \tag{15}$$

The maximization of $\mathcal{W}\left(\mathbf{h}\right)$ gives the conventional delay-and-sum (DS) beamformer [34]:

$$\mathbf{h}_{\text{DS}} = \frac{\mathbf{d}_0}{M}, \tag{16}$$

with $\mathcal{W}\left(\mathbf{h}_{\text{DS}}\right) = M$. While the DS beamformer maximizes the WNG, it never amplifies the diffuse noise. However, $\mathbf{h}_{\text{DS}}$ is not very directional and its beampattern may be very much frequency dependent.

Another important measure, which quantifies the ability of the beamformer to suppress spatial noise from directions other than the endfire direction is the directivity factor (DF), which is obtained by replacing $\mathbf{\Gamma}_{\mathbf{v}}$ with $\mathbf{\Gamma}_{\text{d}}$ in (14), i.e.,

$$\mathcal{D}\left(\mathbf{h}\right) = \frac{\left|\mathbf{h}^H \mathbf{d}_0\right|^2}{\mathbf{h}^H \mathbf{\Gamma}_{\text{d}} \mathbf{h}}. \tag{17}$$

The maximization of $\mathcal{D}\left(\mathbf{h}\right)$ leads to the well-known superdirective beamformer [35]:

$$\mathbf{h}_{\text{SD}} = \frac{\mathbf{\Gamma}_{\text{d}}^{-1} \mathbf{d}_0}{\mathbf{d}_0^H \mathbf{\Gamma}_{\text{d}}^{-1} \mathbf{d}_0}, \tag{18}$$

with $\mathcal{D}\left(\mathbf{h}_{\text{SD}}\right) = \mathbf{d}_0^H \mathbf{\Gamma}_{\text{d}}^{-1} \mathbf{d}_0$. While $\mathbf{h}_{\text{SD}}$ maximizes the DF, it may amplify the white noise, especially at low frequencies, i.e., $\mathcal{W}\left(\mathbf{h}_{\text{SD}}\right) \leq 1$. However, $\mathbf{h}_{\text{SD}}$ is very directional and its beampattern may be frequency invariant.

Finally, the last measure of interest is the beampattern, which describes the sensitivity of the beamformer to a plane wave (source signal) impinging on the array from the direction θ. Mathematically, it is defined as

$$\mathcal{B}_\theta\left(\mathbf{h}\right) = \mathbf{d}_\theta^H \mathbf{h}. \tag{19}$$

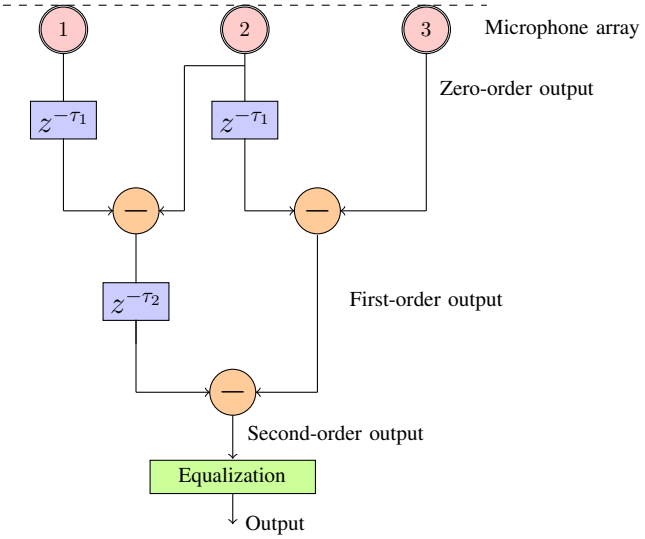

Fig. 1. Schematic diagram of the conventional multistage differential beamforming structure, where the time delays, i.e., τ_1 and τ_2, in each stage are used for controlling the directions of the nulls and forming a beampattern of interest. The frequency and level equalization are used to properly compensate the frequency response of broadband speech signals.

III. Differential Beamforming Theory

Differential microphone arrays refer to those microphone arrays that combine closely spaced sensors to respond to the spatial derivatives of the acoustic pressure field. Early efforts in this area have focused on linear arrays, where differential beamformers are designed in a multistage manner in the time domain, and the differentials of the acoustic pressure field are obtained by combining the outputs of a number of omnidirectional sensors [9], [13], [14], [22]. Figure 1 illustrates how first- and second-order differential microphone arrays are formed. The output of a first-order differential microphone array is the difference between the outputs of two adjacent omnidirectional microphones, and the output of a second-order differential microphone array is the difference between the outputs of two first-order differential microphone arrays. The time delays in each stage are used for controlling the directions of the nulls and forming a beampattern of interest. In such a structure, the response of an Nth-order differential microphone array has a high-pass filter nature with a slope of $6N$ dB/octave. So, its frequency response has to be properly compensated to process broadband speech signals. Moreover, the frequency response and level of a differential microphone array are sensitive to the position and orientation of the arrays relative to the sound source. Consequently, it is necessary to perform frequency and level equalization to its response according to the range and incidence angle of the sound source to obtain the final output [22]. This kind of method apparently lacks flexibility in forming different beampatterns. The resulting differential beamformer is very sensitive to sensor noise and model mismatches, which restricts its practical application.

A different design method of differential beamformers was developed in the short-time Fourier transform (STFT) domain, which involves solving a linear system constructed from null

constraints [16], [17], [25]. In this approach, the robustness of the beamformers can be improved with the minimum-norm solution by increasing the number of microphones [16], [27]. However, with this method, the underlying principle of differential operation in differential microphone arrays becomes less obvious. So, in this section, we present a brief theory of differential beamforming with uniform linear arrays for any order, in which the conventional differential beamforming can be viewed as a particular case.

Let us consider the signal model given in (8). We define the first-order forward spatial difference of $\mathbf{y}$ as

$$\Delta Y_i = Y_{i+1} - Y_i = Y_{(1),i},\ i = 1, 2, \ldots, M-1, \tag{20}$$

where Δ is the forward spatial difference operator. In a vector/matrix form, (20) is

$$\boldsymbol{\Delta}_{(1)}\mathbf{y} = \mathbf{y}_{(1)}, \tag{21}$$

where

$$\boldsymbol{\Delta}_{(1)} = \begin{bmatrix} -1 & 1 & 0 & \cdots & 0 \\ 0 & -1 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & -1 & 1 \end{bmatrix} \tag{22}$$

is a matrix of size $(M-1)\times M$. In the same way, the second-order forward spatial difference of $\mathbf{y}$ is

$$\begin{aligned}\Delta^2 Y_i &= \Delta\left(\Delta Y_i\right) = \Delta Y_{i+1} - \Delta Y_i \\ &= Y_{i+2} - 2Y_{i+1} + Y_i = Y_{(2),i},\ i = 1, 2, \ldots, M-2,\end{aligned} \tag{23}$$

which can be rewritten as

$$\boldsymbol{\Delta}_{(2)}\mathbf{y} = \mathbf{y}_{(2)}, \tag{24}$$

where

$$\boldsymbol{\Delta}_{(2)} = \begin{bmatrix} 1 & -2 & 1 & 0 & \cdots & 0 \\ 0 & 1 & -2 & 1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & -2 & 1 \end{bmatrix} \tag{25}$$

is a matrix of size $(M-2)\times M$. More generally, let $p = 0, 1, \ldots, P$, with $1 \leq P < M$. By definition, we write $\boldsymbol{\Delta}_{(0)} = \mathbf{I}_M$. Therefore,

$$\boldsymbol{\Delta}_{(0)}\mathbf{y} = \mathbf{I}_M\mathbf{y} = \mathbf{y}. \tag{26}$$

We define the pth-order forward spatial difference of $\mathbf{y}$ as

$$\begin{aligned}\Delta^p Y_i &= \Delta^{p-1}\left(\Delta Y_i\right) = \Delta^{p-1}Y_{i+1} - \Delta^{p-1}Y_i \\ &= \sum_{j=0}^{p}(-1)^{p-j}\binom{p}{j} Y_{i+j},\ i = 1, 2, \ldots, M-p,\end{aligned} \tag{27}$$

where

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

is the binomial coefficient. In a vector/matrix form, (27) is

$$\boldsymbol{\Delta}_{(p)}\mathbf{y} = \mathbf{y}_{(p)}, \tag{28}$$

where

$$\boldsymbol{\Delta}_{(p)} = \begin{bmatrix} \mathbf{c}_{(p)}^T & 0 & \cdots & 0 \\ 0 & \mathbf{c}_{(p)}^T & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{c}_{(p)}^T \end{bmatrix} \tag{29}$$

is a matrix of size $(M-p)\times M$, with

$$\begin{aligned}\mathbf{c}_{(p)} = \Bigg[&(-1)^p\binom{p}{0} \quad (-1)^{p-1}\binom{p}{1} \\ &\cdots \quad -\binom{p}{p-1} \quad 1\Bigg]^T\end{aligned} \tag{30}$$

being a vector of length $p+1$.

Now, substituting (8) into (27), we get

$$\begin{aligned}\Delta^p Y_i &= X\Delta^p D_{0,i} + \Delta^p V_i \\ &= X\sum_{j=0}^{p}(-1)^{p-j}\binom{p}{j} D_{0,i+j} + \Delta^p V_i \\ &= \tau_0^p D_{0,i} X + \Delta^p V_i,\ i = 1, 2, \ldots, M-p,\end{aligned} \tag{31}$$

where

$$\begin{aligned}\tau_0^p &= \sum_{j=0}^{p}(-1)^{p-j}\binom{p}{j} D_{0,j+1} \\ &= \sum_{j=0}^{p}(-1)^{p-j}\binom{p}{j} e^{-\jmath\varpi_0 j} \\ &= \left(e^{-\jmath\varpi_0} - 1\right)^p.\end{aligned} \tag{32}$$

In a vector/matrix form, (31) becomes

$$\boldsymbol{\Delta}_{(p)}\mathbf{y} = \tau_0^p\mathbf{d}_{0,M-p}X + \mathbf{v}_{(p)} = \mathbf{y}_{(p)}, \tag{33}$$

where

$$\mathbf{d}_{0,M-p} = \begin{bmatrix} 1 & e^{-\jmath\varpi_0} & \cdots & e^{-\jmath(M-p-1)\varpi_0} \end{bmatrix}^T \tag{34}$$

is the steering vector of length $M-p$ at $\theta = 0$ and $\mathbf{v}_{(p)} = \boldsymbol{\Delta}_{(p)}\mathbf{v}$. We deduce that the $(M-p)\times(M-p)$ covariance matrix of $\mathbf{y}_{(p)}$ is

$$\begin{aligned}\boldsymbol{\Phi}_{\mathbf{y}_{(p)}} &= \phi_X\left|\tau_0\right|^{2p}\mathbf{d}_{0,M-p}\mathbf{d}_{0,M-p}^H + \boldsymbol{\Delta}_{(p)}\boldsymbol{\Phi}_{\mathbf{v}}\boldsymbol{\Delta}_{(p)}^T \\ &= \phi_X\left|\tau_0\right|^{2p}\mathbf{d}_{0,M-p}\mathbf{d}_{0,M-p}^H + \phi_V\boldsymbol{\Delta}_{(p)}\boldsymbol{\Gamma}_{\mathbf{v}}\boldsymbol{\Delta}_{(p)}^T.\end{aligned} \tag{35}$$

With diffuse noise, (35) is

$$\boldsymbol{\Phi}_{\mathbf{y}_{(p)}} = \phi_X\left|\tau_0\right|^{2p}\mathbf{d}_{0,M-p}\mathbf{d}_{0,M-p}^H + \phi_{\mathrm{d}}\boldsymbol{\Delta}_{(p)}\boldsymbol{\Gamma}_{\mathrm{d}}\boldsymbol{\Delta}_{(p)}^T. \tag{36}$$

Then, from (35), we find that the SNR corresponding to $\mathbf{y}_{(p)}$ is

$$\begin{aligned}\mathrm{SNR}_{\mathbf{y}_{(p)}} &= \frac{\mathrm{tr}\left(\phi_X\left|\tau_0\right|^{2p}\mathbf{d}_{0,M-p}\mathbf{d}_{0,M-p}^H\right)}{\mathrm{tr}\left(\phi_V\boldsymbol{\Delta}_{(p)}\boldsymbol{\Gamma}_{\mathbf{v}}\boldsymbol{\Delta}_{(p)}^T\right)} \\ &= \mathrm{SNR}_{\mathbf{y}} \times \frac{\left|\tau_0\right|^{2p}(M-p)}{\mathrm{tr}\left(\boldsymbol{\Delta}_{(p)}\boldsymbol{\Gamma}_{\mathbf{v}}\boldsymbol{\Delta}_{(p)}^T\right)}.\end{aligned} \tag{37}$$

Therefore, we deduce that the SNR gain (between $\mathbf{y}_{(p)}$ and $\mathbf{y}$) is

$$\mathcal{G}_{(p)} = \frac{\mathrm{SNR}_{\mathbf{y}_{(p)}}}{\mathrm{SNR}_{\mathbf{y}}} = \frac{|\tau_0|^{2p}(M-p)}{\mathrm{tr}\left(\mathbf{\Delta}_{(p)}\mathbf{\Gamma}_{\mathbf{v}}\mathbf{\Delta}_{(p)}^T\right)}. \tag{38}$$

For white and diffuse noises, $\mathcal{G}_{(p)}$ becomes

$$\begin{aligned}\mathcal{W}_{(p)} &= \frac{|\tau_0|^{2p}(M-p)}{\mathrm{tr}\left(\mathbf{\Delta}_{(p)}\mathbf{\Delta}_{(p)}^T\right)} = \frac{|\tau_0|^{2p}(M-p)}{(M-p)\mathbf{c}_{(p)}^T\mathbf{c}_{(p)}} \\ &= \frac{|\tau_0|^{2p}}{\mathbf{c}_{(p)}^T\mathbf{c}_{(p)}}\end{aligned} \tag{39}$$

and

$$\mathcal{D}_{(p)} = \frac{|\tau_0|^{2p}(M-p)}{\mathrm{tr}\left(\mathbf{\Delta}_{(p)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(p)}^T\right)}, \tag{40}$$

respectively, where $\mathbf{c}_{(p)}$ is defined in (30) and it is easy to verify that

$$\mathbf{c}_{(p)}^T\mathbf{c}_{(p)} = \begin{pmatrix} 2p \\ p \end{pmatrix}.$$

One can check that

$$\mathcal{W}_{(0)} = 1 \geq \mathcal{W}_{(1)} \geq \mathcal{W}_{(2)} \geq \cdots \geq \mathcal{W}_{(M-1)}, \tag{41}$$

$$\mathcal{D}_{(0)} = 1 \leq \mathcal{D}_{(1)} \leq \mathcal{D}_{(2)} \leq \cdots \leq \mathcal{D}_{(M-1)}. \tag{42}$$

We conclude that $\mathbf{y}_{(p)}$ may be good to improve diffuse noise but will certainly amplify white noise. In practice, it is paramount to find a good compromise between these two types of noises.

Let us assume that we want to design a Pth-order differential beamformer with M omnidirectional microphones, where $P < M$. For that, we need to apply a complex-valued linear filter, $\mathbf{h}_{(P)}$ of length $M-P$, to the differential observed signal $\mathbf{y}_{(P)}$, i.e.,

$$Z_{(P)} = \mathbf{h}_{(P)}^H\mathbf{y}_{(P)} = X_{\mathrm{fd},(P)} + V_{\mathrm{rn},(P)}, \tag{43}$$

where $Z_{(P)}$ is the differential beamformer output signal (i.e., the estimate of X), $X_{\mathrm{fd},(P)} = X\tau_0^P\mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P}$ is the filtered desired signal, and $V_{\mathrm{rn},(P)} = \mathbf{h}_{(P)}^H\mathbf{v}_{(P)}$ is the residual noise. As a result, the variance of $Z_{(P)}$ is

$$\phi_{Z_{(P)}} = \mathbf{h}_{(P)}^H\mathbf{\Phi}_{\mathbf{y}_{(P)}}\mathbf{h}_{(P)} = \phi_{X_{\mathrm{fd},(P)}} + \phi_{V_{\mathrm{rn},(P)}}, \tag{44}$$

where $\phi_{X_{\mathrm{fd},(P)}} = \phi_X|\tau_0|^{2P}\left|\mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P}\right|^2$ and $\phi_{V_{\mathrm{rn},(P)}} = \mathbf{h}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Phi}_{\mathbf{v}}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P)}$. Also, it is clear that the distortionless constraint is

$$\mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P} = \frac{1}{\tau_0^P}. \tag{45}$$

Now, let us briefly give the most important performance measures corresponding to $\mathbf{h}_{(P)}$. The output SNR and the gain in SNR are, respectively,

$$\mathrm{oSNR}\left(\mathbf{h}_{(P)}\right) = \frac{\phi_X}{\phi_V} \times \frac{|\tau_0|^{2P}\left|\mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P}\right|^2}{\mathbf{h}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathbf{v}}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P)}} \tag{46}$$

and

$$\mathcal{G}\left(\mathbf{h}_{(P)}\right) = \frac{\mathrm{oSNR}\left(\mathbf{h}_{(P)}\right)}{\mathrm{iSNR}} = \frac{|\tau_0|^{2P}\left|\mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P}\right|^2}{\mathbf{h}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathbf{v}}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P)}}, \tag{47}$$

from which we deduce the WNG:

$$\mathcal{W}\left(\mathbf{h}_{(P)}\right) = \frac{|\tau_0|^{2P}\left|\mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P}\right|^2}{\mathbf{h}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P)}} \tag{48}$$

and the DF:

$$\mathcal{D}\left(\mathbf{h}_{(P)}\right) = \frac{|\tau_0|^{2P}\left|\mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P}\right|^2}{\mathbf{h}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P)}}. \tag{49}$$

Finally, the power beampattern is

$$\left|\mathcal{B}_\theta\left(\mathbf{h}_{(P)}\right)\right|^2 = |\tau_0|^{2P}\left|\mathbf{h}_{(P)}^H\mathbf{d}_{\theta,M-P}\right|^2. \tag{50}$$

IV. Optimal Differential Beamformers

Now, we are ready to develop some obvious and useful optimal Pth-order differential beamformers from the above proposed theory. The focus is on the derivation of fixed beamformers only; but the extension to adaptive beamformers is also possible.

A. Maximum WNG/DF Differential Beamformers

The first beamformer is deduced from the maximization of the WNG in (48), i.e.,

$$\begin{aligned}&\min_{\mathbf{h}_{(P)}} \mathbf{h}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P)} \\ &\quad \text{subject to} \quad \mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P} = \frac{1}{\tau_0^P}.\end{aligned} \tag{51}$$

From the previous minimization, we get the Pth-order maximum WNG (MWNG) differential beamformer:

$$\mathbf{h}_{(P),\mathrm{MWNG}} = \frac{\left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}}{\left(\tau_0^*\right)^P\mathbf{d}_{0,M-P}^H\left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}}, \tag{52}$$

where the superscript * is the complex-conjugate operator.

The second beamformer of interest is the one derived from the maximization of the DF in (49). This maximization is equivalent to

$$\begin{aligned}&\min_{\mathbf{h}_{(P)}} \mathbf{h}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P)} \\ &\quad \text{subject to} \quad \mathbf{h}_{(P)}^H\mathbf{d}_{0,M-P} = \frac{1}{\tau_0^P},\end{aligned} \tag{53}$$

from which we deduce the Pth-order maximum DF (MDF) differential beamformer:

$$\mathbf{h}_{(P),\mathrm{MDF}} = \frac{\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}}{\left(\tau_0^*\right)^P\mathbf{d}_{0,M-P}^H\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}}. \tag{54}$$

B. Parameterized MDF Differential Beamformers

The MDF beamformer will lead to a high value of the DF but at the expense of some white noise amplification, while the MWNG beamformer will lead to a large value of the WNG but at the expense of some low value of the DF. Therefore, in practice, it is important to be able to compromise between WNG and DF. To do so, we propose to exploit the joint diagonalization technique of two Hermitian matrices (different versions of this technique have been used in many applications, e.g., noise reduction and beamforming [36]–[38].

The two Hermitian matrices $|\tau_0|^{2P}\,\mathbf{d}_{0,M-P}\mathbf{d}_{0,M-P}^H$ and $\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T$, which appear in the definition of the DF, can be jointly diagonalized as follows [39]:

$$|\tau_0|^{2P}\,\mathbf{T}_{(P)}^H\mathbf{d}_{0,M-P}\mathbf{d}_{0,M-P}^H\mathbf{T}_{(P)} = \mathbf{\Lambda}_{(P)}, \tag{55}$$

$$\mathbf{T}_{(P)}^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\mathbf{T}_{(P)} = \mathbf{I}_{M-P}, \tag{56}$$

where

$$\mathbf{T}_{(P)} = \begin{bmatrix} \mathbf{t}_{(P),1} & \mathbf{t}_{(P),2} & \cdots & \mathbf{t}_{(P),M-P} \end{bmatrix} \tag{57}$$

is a full-rank square matrix of size $(M-P)\times(M-P)$,

$$\mathbf{t}_{(P),1} = \frac{\tau_0^P\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}}{\sqrt{|\tau_0|^{2P}\,\mathbf{d}_{0,M-P}^H\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}}} \tag{58}$$

is the first eigenvector of the matrix $|\tau_0|^{2P}\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}\mathbf{d}_{0,M-P}^H$,

$$\mathbf{\Lambda}_{(P)} = \mathrm{diag}\left(\lambda_{(P),1},0,\ldots,0\right) \tag{59}$$

is a diagonal matrix of size $(M-P)\times(M-P)$, $\mathbf{I}_{M-P}$ is the $(M-P)\times(M-P)$ identity matrix, and

$$\lambda_{(P),1} = |\tau_0|^{2P}\,\mathbf{d}_{0,M-P}^H\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P} \tag{60}$$

is the only nonnull eigenvalue of $|\tau_0|^{2P}\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}\mathbf{d}_{0,M-P}^H$, whose corresponding eigenvector is $\mathbf{t}_{(P),1}$. It can be checked from (55) that

$$\mathbf{t}_{(P),i}^H\mathbf{d}_{0,M-P} = 0,\ i = 2,3,\ldots,M-P. \tag{61}$$

Let us define the matrix of size $(M-P)\times N$:

$$\mathbf{T}_{(P),1:N} = \begin{bmatrix} \mathbf{t}_{(P),1} & \mathbf{t}_{(P),2} & \cdots & \mathbf{t}_{(P),N} \end{bmatrix}, \tag{62}$$

with $1\leq N\leq M-P$. The differential beamformer that we consider now has the form:

$$\mathbf{h}_{(P),1:N} = \mathbf{T}_{(P),1:N}\boldsymbol{\alpha}, \tag{63}$$

where

$$\boldsymbol{\alpha} = \begin{bmatrix} \alpha_1 & \alpha_2 & \cdots & \alpha_N \end{bmatrix}^T \neq \mathbf{0} \tag{64}$$

is a vector of length N. Substituting (63) into (43), we find that

$$\begin{aligned} Z_{(P)} &= \tau_0^P\boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{d}_{0,M-P}X + \boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{v} \\ &= \alpha_1^*\sqrt{\lambda_{(P),1}}X + \boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{v}. \end{aligned} \tag{65}$$

Since the distortionless constraint is desired, it is clear from the previous expression that we always choose

$$\alpha_1 = \frac{1}{\sqrt{\lambda_{(1),1}}}. \tag{66}$$

Now, we need to determine the other elements of $\boldsymbol{\alpha}$.

Thanks to the nature of the beamformer given in (63), we can express the WNG and the DF, respectively, as

$$\begin{aligned} \mathcal{W}\left(\mathbf{h}_{(P),1:N}\right) &= \frac{|\tau_0|^{2P}\left|\mathbf{h}_{(P),1:N}^H\mathbf{d}_{0,M-P}\right|^2}{\mathbf{h}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P),1:N}} \\ &= \frac{|\tau_0|^{2P}\left|\boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{d}_{0,M-P}\right|^2}{\boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\mathbf{T}_{(P),1:N}\boldsymbol{\alpha}} \end{aligned} \tag{67}$$

and

$$\begin{aligned} \mathcal{D}\left(\mathbf{h}_{(P),1:N}\right) &= \frac{|\tau_0|^{2P}\left|\mathbf{h}_{(P),1:N}^H\mathbf{d}_{0,M-P}\right|^2}{\mathbf{h}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\mathbf{h}_{(P),1:N}} \\ &= \frac{|\tau_0|^{2P}\left|\boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{d}_{0,M-P}\right|^2}{\boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\mathbf{T}_{(P),1:N}\boldsymbol{\alpha}} \\ &= \frac{|\tau_0|^{2P}\left|\boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{d}_{0,M-P}\right|^2}{\boldsymbol{\alpha}^H\boldsymbol{\alpha}}. \end{aligned} \tag{68}$$

Consequently, if we want to compromise between WNG and DF, we need to maximize the WNG given above, i.e.,

$$\begin{aligned} &\min_{\boldsymbol{\alpha}}\ \boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\mathbf{T}_{(P),1:N}\boldsymbol{\alpha} \\ &\quad\text{subject to}\quad \boldsymbol{\alpha}^H\mathbf{T}_{(P),1:N}^H\mathbf{d}_{0,M-P} = \frac{1}{\tau_0^P}. \end{aligned} \tag{69}$$

Substituting the obtained $\boldsymbol{\alpha}$ from (69) into (63), we easily deduce the parameterized MDF beamformer:

$$\mathbf{h}_{(P),1:N} = \frac{\mathbf{P}_{\mathbf{T}_{(P),1:N}}\mathbf{d}_{0,M-P}}{\left(\tau_0^*\right)^P\mathbf{d}_{0,M-P}^H\mathbf{P}_{\mathbf{T}_{(P),1:N}}\mathbf{d}_{0,M-P}}, \tag{70}$$

where

$$\begin{aligned} &\mathbf{P}_{\mathbf{T}_{(P),1:N}} = \\ &\mathbf{T}_{(P),1:N}\left(\mathbf{T}_{(P),1:N}^H\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\mathbf{T}_{(P),1:N}\right)^{-1}\mathbf{T}_{(P),1:N}^H. \end{aligned} \tag{71}$$

For $N=1$, we get

$$\mathbf{h}_{(P),1:1} = \frac{\mathbf{t}_{(P),1}}{\left(\tau_0^*\right)^P\mathbf{d}_{0,M-P}^H\mathbf{t}_{(P),1}} = \mathbf{h}_{(P),\mathrm{MDF}}, \tag{72}$$

which is the Pth-order MDF differential beamformer, and for $N=M-P$, we obtain

$$\begin{aligned} \mathbf{h}_{(P),1:M-P} &= \frac{\left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}}{\left(\tau_0^*\right)^P\mathbf{d}_{0,M-P}^H\left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{d}_{0,M-P}} \\ &= \mathbf{h}_{(P),\mathrm{MWNG}}, \end{aligned} \tag{73}$$

which is the Pth-order MWNG differential beamformer. Therefore, by adjusting the positive integer N, we can obtain different beamformers whose performances are in between the

performances of $\mathbf{h}_{(P),\mathrm{MDF}}$ and $\mathbf{h}_{(P),\mathrm{MWNG}}$. Note that there are some other ways to derive parameterized beamformers, e.g, via linear interpolation, whose performance is a tradeoff between those two beamformers [40].

C. Parameterized Maximum Front-to-Back Ratio Differential Beamformers

Another important performance measure that we have not discussed so far but is useful in the context of uniform linear arrays is the front-to-back ratio [41], 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. In the conventional beamforming case, it is given by

$$\mathcal{F}\left(\mathbf{h}\right)=\frac{\displaystyle\int_{0}^{\pi/2}\left|\mathcal{B}_{\theta}\left(\mathbf{h}\right)\right|^{2}\sin\theta d\theta}{\displaystyle\int_{\pi/2}^{\pi}\left|\mathcal{B}_{\theta}\left(\mathbf{h}\right)\right|^{2}\sin\theta d\theta}=\frac{\mathbf{h}^{H}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{h}}{\mathbf{h}^{H}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{h}},\tag{74}$$

where

$$\mathbf{\Gamma}_{\mathrm{f}}=\int_{0}^{\pi/2}\mathbf{d}_{\theta}\mathbf{d}_{\theta}^{H}\sin\theta d\theta,\tag{75}$$

$$\mathbf{\Gamma}_{\mathrm{b}}=\int_{\pi/2}^{\pi}\mathbf{d}_{\theta}\mathbf{d}_{\theta}^{H}\sin\theta d\theta.\tag{76}$$

Therefore, the front-to-back ratio of the Pth-order differential beamformer is

$$\begin{aligned}\mathcal{F}\left(\mathbf{h}_{(P)}\right)&=\frac{\displaystyle\int_{0}^{\pi/2}\left|\mathcal{B}_{\theta}\left(\mathbf{h}_{(P)}\right)\right|^{2}\sin\theta d\theta}{\displaystyle\int_{\pi/2}^{\pi}\left|\mathcal{B}_{\theta}\left(\mathbf{h}_{(P)}\right)\right|^{2}\sin\theta d\theta}\\&=\frac{\mathbf{h}_{(P)}^{H}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{\Delta}_{(P)}^{T}\mathbf{h}_{(P)}}{\mathbf{h}_{(P)}^{H}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{\Delta}_{(P)}^{T}\mathbf{h}_{(P)}}.\end{aligned}\tag{77}$$

Let us denote by $\widetilde{\mathbf{t}}_{(P),1}$ the eigenvector corresponding to the maximum eigenvalue of the matrix $\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{\Delta}_{(P)}^{T}\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{\Delta}_{(P)}^{T}$. Taking into account the distortionless constraint, we easily get the Pth-order maximum front-to-back ratio differential beamformer:

$$\mathbf{h}_{(P),\mathrm{MFBR}}=\frac{\widetilde{\mathbf{t}}_{(P),1}}{\mathbf{d}_{0}^{H}\mathbf{\Delta}_{(P)}^{T}\widetilde{\mathbf{t}}_{(P),1}},\tag{78}$$

where we recall that $\mathbf{\Delta}_{(P)}\mathbf{d}_{0}=\tau_{0}^{P}\mathbf{d}_{0,M-P}$. While this beamformer maximizes the front-to-back ratio in (77), it will likely amplify the white noise.

In order to better compromise between front-to-back ratio and WNG, we propose to jointly diagonalize the two Hermitian matrices $\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{\Delta}_{(P)}^{T}$ and $\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{\Delta}_{(P)}^{T}$ that appear in the front-to-back ratio. We have [39]

$$\widetilde{\mathbf{T}}_{(P)}^{H}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{\Delta}_{(P)}^{T}\widetilde{\mathbf{T}}_{(P)}=\widetilde{\mathbf{\Lambda}}_{(P)},\tag{79}$$

$$\widetilde{\mathbf{T}}_{(P)}^{H}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{\Delta}_{(P)}^{T}\widetilde{\mathbf{T}}_{(P)}=\mathbf{I}_{M-P},\tag{80}$$

where

$$\widetilde{\mathbf{T}}_{(P)}=\left[\begin{array}{cccc}\widetilde{\mathbf{t}}_{(P),1} & \widetilde{\mathbf{t}}_{(P),2} & \cdots & \widetilde{\mathbf{t}}_{(P),M-P}\end{array}\right]\tag{81}$$

is a full-rank square matrix [of size $(M-P)\times(M-P)$] and

$$\widetilde{\mathbf{\Lambda}}_{(P)}=\mathrm{diag}\left(\widetilde{\lambda}_{(P),1},\widetilde{\lambda}_{(P),2},\ldots,\widetilde{\lambda}_{(P),M-P}\right)\tag{82}$$

is a diagonal matrix whose main elements are real and nonnegative. Furthermore, $\widetilde{\mathbf{\Lambda}}_{(P)}$ and $\widetilde{\mathbf{T}}_{(P)}$ are the eigenvalue and eigenvector matrices, respectively, of $\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{\Delta}_{(P)}^{T}\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{\Delta}_{(P)}^{T}$, i.e.,

$$\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{\Delta}_{(P)}^{T}\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{\Delta}_{(P)}^{T}\widetilde{\mathbf{T}}_{(P)}=\widetilde{\mathbf{T}}_{(P)}\widetilde{\mathbf{\Lambda}}_{(P)}.\tag{83}$$

It is assumed that the eigenvalues of $\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{b}}\mathbf{\Delta}_{(P)}^{T}\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{f}}\mathbf{\Delta}_{(P)}^{T}$ are ordered as $\widetilde{\lambda}_{(P),1}\geq\widetilde{\lambda}_{(P),2}\geq\cdots\geq\widetilde{\lambda}_{(P),M-P}\geq0$. Therefore, the corresponding eigenvectors are $\widetilde{\mathbf{t}}_{(P),1},\widetilde{\mathbf{t}}_{(P),2},\ldots,\widetilde{\mathbf{t}}_{(P),M-P}$.

Let us define the matrix of size $(M-P)\times N$:

$$\widetilde{\mathbf{T}}_{(P),1:N}=\left[\begin{array}{cccc}\widetilde{\mathbf{t}}_{(P),1} & \widetilde{\mathbf{t}}_{(P),2} & \cdots & \widetilde{\mathbf{t}}_{(P),N}\end{array}\right],\tag{84}$$

with $1\leq N\leq M-P$. We consider the differential beamformer whose form is

$$\widetilde{\mathbf{h}}_{(P),1:N}=\widetilde{\mathbf{T}}_{(P),1:N}\widetilde{\boldsymbol{\alpha}},\tag{85}$$

where

$$\widetilde{\boldsymbol{\alpha}}=\left[\begin{array}{cccc}\widetilde{\alpha}_{1} & \widetilde{\alpha}_{2} & \cdots & \widetilde{\alpha}_{N}\end{array}\right]^{T}\neq\mathbf{0}\tag{86}$$

is a vector of length N.

Now, we propose to optimize the criterion:

$$\begin{aligned}&\min_{\widetilde{\boldsymbol{\alpha}}}\widetilde{\boldsymbol{\alpha}}^{H}\widetilde{\mathbf{T}}_{(P),1:N}^{H}\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^{T}\widetilde{\mathbf{T}}_{(P),1:N}\widetilde{\boldsymbol{\alpha}}\\&\quad\text{subject to}\quad\widetilde{\boldsymbol{\alpha}}^{H}\widetilde{\mathbf{T}}_{(P),1:N}^{H}\mathbf{d}_{0,M-P}=\frac{1}{\tau_{0}^{P}}.\end{aligned}\tag{87}$$

Substituting the obtained $\widetilde{\boldsymbol{\alpha}}$ from (87) into (85) gives the parameterized maximum front-to-back ratio beamformer:

$$\widetilde{\mathbf{h}}_{(P),1:N}=\frac{\mathbf{P}_{\widetilde{\mathbf{T}}_{(P),1:N}}\mathbf{d}_{0,M-P}}{\left(\tau_{0}^{*}\right)^{P}\mathbf{d}_{0,M-P}^{H}\mathbf{P}_{\widetilde{\mathbf{T}}_{(P),1:N}}\mathbf{d}_{0,M-P}},\tag{88}$$

where

$$\begin{aligned}&\mathbf{P}_{\widetilde{\mathbf{T}}_{(P),1:N}}=\\&\widetilde{\mathbf{T}}_{(P),1:N}\left(\widetilde{\mathbf{T}}_{(P),1:N}^{H}\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^{T}\widetilde{\mathbf{T}}_{(P),1:N}\right)^{-1}\widetilde{\mathbf{T}}_{(P),1:N}^{H}.\end{aligned}\tag{89}$$

For $N=1$, we get

$$\begin{aligned}\widetilde{\mathbf{h}}_{(P),1:1}&=\frac{\widetilde{\mathbf{t}}_{(P),1}}{\left(\tau_{0}^{*}\right)^{P}\mathbf{d}_{0,M-P}^{H}\widetilde{\mathbf{t}}_{(P),1}}\\&=\mathbf{h}_{(P),\mathrm{MFBR}},\end{aligned}\tag{90}$$

which is the Pth-order maximum front-to-back ratio differential beamformer, and for $N=M-P$, we obtain

$$\begin{aligned}\widetilde{\mathbf{h}}_{(P),1:M-P}&=\frac{\left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^{T}\right)^{-1}\mathbf{d}_{0,M-P}}{\left(\tau_{0}^{*}\right)^{P}\mathbf{d}_{0,M-P}^{H}\left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^{T}\right)^{-1}\mathbf{d}_{0,M-P}}\\&=\mathbf{h}_{(P),\mathrm{MWNG}},\end{aligned}\tag{91}$$

which is the Pth-order MWNG differential beamformer. Therefore, by adjusting the positive integer N, we can obtain different beamformers whose performances are in between the performances of $\mathbf{h}_{(P),\mathrm{MFBR}}$ and $\mathbf{h}_{(P),\mathrm{MWNG}}$.

D. Null-Constraint Differential Beamformers

In practice, in order to give a desired shape to the pattern beamformer, a null constraint in the direction $\theta_0 \in [\pi/2, \pi]$ may be needed. Considering also the distortionless constraint, we can build the constraint equation as

$$\mathbf{C}^H \mathbf{\Delta}_{(P)}^T \mathbf{h}_{(P)} = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \tag{92}$$

where

$$\mathbf{C} = \begin{bmatrix} \mathbf{d}_0 & \mathbf{d}_{\theta_0} \end{bmatrix} \tag{93}$$

is the constraint matrix of size $M \times 2$ whose columns are linearly independent. We recall that we have

$$\mathbf{\Delta}_{(P)}\mathbf{C} = \begin{bmatrix} \tau_0^P \mathbf{d}_{0,M-P} & \tau_{\theta_0}^P \mathbf{d}_{\theta_0,M-P} \end{bmatrix}.$$

Therefore, by maximizing the WNG, we obtain the Pth-order null-constraint MWNG (NCMWNG) differential beamformer:

$$\mathbf{h}_{(P),\mathrm{NCMWNG}} = \left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{C}\times \left[\mathbf{C}^H\mathbf{\Delta}_{(P)}^T\left(\mathbf{\Delta}_{(P)}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{C}\right]^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix} \tag{94}$$

and by maximizing the DF, we obtain the Pth-order null-constraint MDF (NCMDF) differential beamformer:

$$\mathbf{h}_{(P),\mathrm{NCMDF}} = \left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{C}\times \left[\mathbf{C}^H\mathbf{\Delta}_{(P)}^T\left(\mathbf{\Delta}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\mathbf{\Delta}_{(P)}^T\right)^{-1}\mathbf{\Delta}_{(P)}\mathbf{C}\right]^{-1}\begin{bmatrix} 1 \\ 0 \end{bmatrix}. \tag{95}$$

E. Combined Differential Beamformers

Now, let us briefly discuss another interesting perspective. Since M microphones are available, the number of degrees of freedom is also equal to M. However, the dimension of $\mathbf{y}_{(P)}$ is only equal to $M - P$. Therefore, we can add P more observations to $\mathbf{y}_{(P)}$. For example, by taking the first P components of $\mathbf{y}$, we may consider the observed signal of length M:

$$\begin{aligned}\overrightarrow{\mathbf{y}} &= \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_P & \mathbf{y}_{(P)}^T \end{bmatrix}^T \\ &= \begin{bmatrix} \mathbf{y}_P^T & \mathbf{y}_{(P)}^T \end{bmatrix}^T \\ &= \overrightarrow{\mathbf{d}}_0 X + \overrightarrow{\mathbf{v}},\end{aligned} \tag{96}$$

which is a combination of pressure and difference pressure observations, where

$$\overrightarrow{\mathbf{d}}_0 = \begin{bmatrix} \mathbf{d}_{0,P}^T & \tau_0^P \mathbf{d}_{0,M-P}^T \end{bmatrix}^T$$

and $\overrightarrow{\mathbf{v}}$ is defined similarly to $\overrightarrow{\mathbf{y}}$. Consequently, the proposed beamformer output is

$$Z_{\overrightarrow{\mathbf{h}}} = \overrightarrow{\mathbf{h}}^H \overrightarrow{\mathbf{y}}, \tag{97}$$

where $\overrightarrow{\mathbf{h}}$ is a beamformer of length M. Obviously, many optimal beamformers can be derived with this approach as suggested above. For examples, we easily derive the MWNG beamformer:

$$\overrightarrow{\mathbf{h}}_{\mathrm{MWNG}} = \frac{\left(\overrightarrow{\mathbf{\Delta}}_{(P)}\overrightarrow{\mathbf{\Delta}}_{(P)}^T\right)^{-1}\overrightarrow{\mathbf{d}}_0}{\overrightarrow{\mathbf{d}}_0^H\left(\overrightarrow{\mathbf{\Delta}}_{(P)}\overrightarrow{\mathbf{\Delta}}_{(P)}^T\right)^{-1}\overrightarrow{\mathbf{d}}_0} \tag{98}$$

and the MDF beamformer:

$$\overrightarrow{\mathbf{h}}_{\mathrm{MDF}} = \frac{\left(\overrightarrow{\mathbf{\Delta}}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\overrightarrow{\mathbf{\Delta}}_{(P)}^T\right)^{-1}\overrightarrow{\mathbf{d}}_0}{\overrightarrow{\mathbf{d}}_0^H\left(\overrightarrow{\mathbf{\Delta}}_{(P)}\mathbf{\Gamma}_{\mathrm{d}}\overrightarrow{\mathbf{\Delta}}_{(P)}^T\right)^{-1}\overrightarrow{\mathbf{d}}_0}, \tag{99}$$

where

$$\overrightarrow{\mathbf{\Delta}}_{(P)} = \begin{bmatrix} \mathbf{I}_P & \mathbf{0} \\ & \mathbf{\Delta}_{(P)} \end{bmatrix}, \tag{100}$$

with $\mathbf{I}_P$ being the $P \times P$ identity matrix.

While we deduced the differential beamforming in two steps, they can easily be combined into a single one as

$$\mathbf{h} = \mathbf{\Delta}_{(p)}^T \mathbf{h}_{(P)}, \tag{101}$$

since

$$\begin{aligned}Z_{(P)} &= \mathbf{h}_{(P)}^H \mathbf{y}_{(P)} = \mathbf{h}_{(P)}^H \mathbf{\Delta}_{(p)}\mathbf{y} = (\mathbf{\Delta}_{(p)}^T\mathbf{h}_{(P)})^H\mathbf{y} \\ &= \mathbf{h}^H\mathbf{y}.\end{aligned} \tag{102}$$

V. Simulations

A. Performance Study

In this section, we study the performance of the developed differential beamformers in terms of the beampattern, WNG, and directivity index (DI), where the DI is the DF expressed in decibels [9], i.e., $\mathrm{DI}(\mathbf{h}) = 10\log_{10}\mathcal{D}(\mathbf{h})$. We consider a uniform linear array with an interelement spacing of 1.0 cm. The desired source signal propagates from the endfire direction, i.e., $\theta = 0^\circ$.

We first study the performance of the basic forward spatial differential operator, where the WNG and DF are computed according to (39) and (40), respectively. Figure 2 shows plots of the DI and the WNG of this operator for $p = 1, 2, 3$, as a function of frequency, f. It is clearly seen that this operator achieves significant directivity gains and the DIs are frequency invariant. We see that the DI increases while the WNG decreases as p increases. The WNG is considerably less than 0 dB, which indicates that the forward spatial differential operator suffers from significant white noise amplification. This operator serves as the first stage of the beamforming process, so its high directional and frequency-invariant properties also delineate the essence and advantage of differential beamforming.

Figure 3 plots the DI and the WNG of the parameterized MDF differential beamformer with $M = 6$ and $P = 1$, as a function of frequency, f, for different values of N. In this simulation, we use six microphones to design a first-order differential beamformer, so that $N = 1, 2, 3, 5$. As seen from Fig. 3, the DI decreases while the WNG increases with N. Clearly, we can choose a proper value of N to design

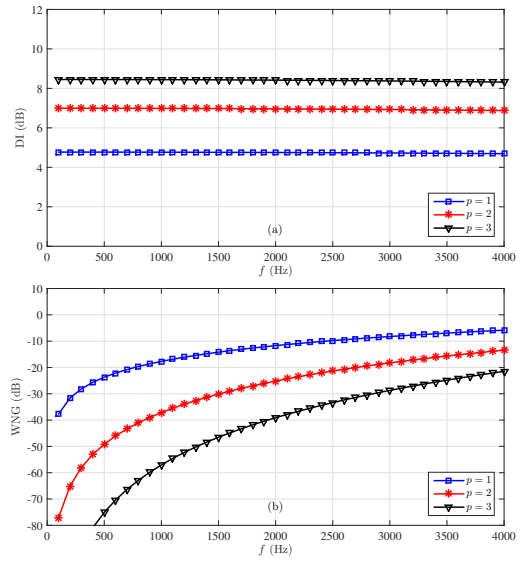

Fig. 2. Performance of the forward spatial differential operator as a function of frequency for different orders, p: (a) DI and (b) WNG. Conditions of simulation: $M = 4$ and $\delta = 1.0$ cm.

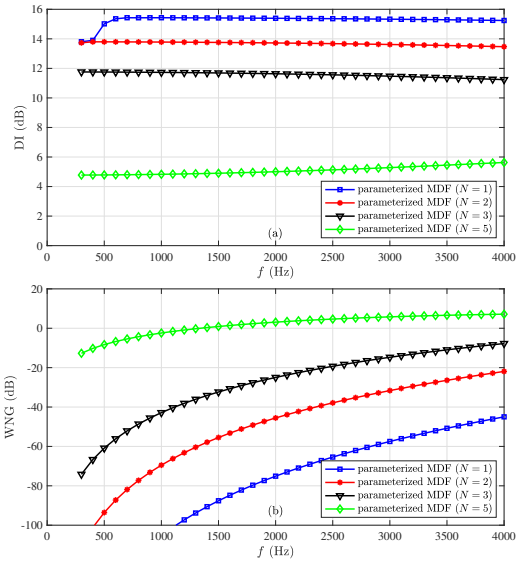

Fig. 3. Performance of the parameterized MDF differential beamformer as a function of frequency for different values of N: (a) DI and (b) WNG. Conditions of simulation: $M = 6$, $\delta = 1.0$ cm, and $P = 1$.

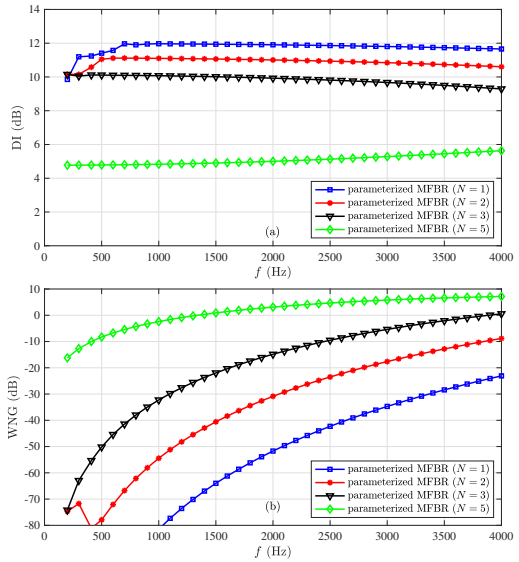

Fig. 4. Performance of the parameterized MFBR differential beamformer as a function of frequency for different values of N: (a) DI and (b) WNG. Conditions of simulation: $M = 6$, $\delta = 1.0$ cm, and $P = 1$.

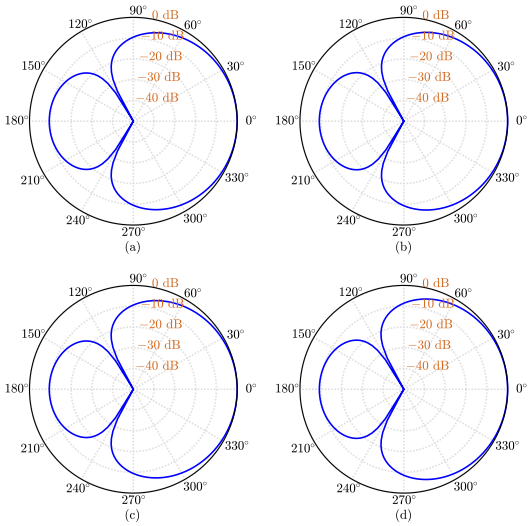

Fig. 5. Beampatterns of the NCMWNG differential beamformer for different numbers of microphones: (a) $M = 4$, $f = 1000$ Hz, (b) $M = 4$, $f = 3000$ Hz, (c) $M = 6$, $f = 1000$ Hz, and (d) $M = 6$, $f = 3000$ Hz. Conditions: $\delta = 1.0$ cm and $P = 2$.

a differential beamformer that achieves a good compromise between a large value of the DI and white noise amplification.

Figure 4 plots the DI and the WNG of the parameterized maximum front-to-back ratio (MFBR) beamformer with $M = 6$ and $P = 1$, as a function of frequency, f, for different values of N. Similarly, it is clearly seen the DI decreases while the WNG increases with N, and a proper value of N can make the differential beamformer achieve a good compromise among the DI, FBR, and WNG.

We then study the performance of the NCMWNG differential beamformers. In this simulation, we design a second-order differential beamformer, i.e., $P = 2$, and add a null constraint at 120°. Figure 5 shows plots of the beampatterns of the NCMWNG beamformer at different frequencies for $M = 4, 6, 8, 12$. It can be observed that the designed beampatterns are almost frequency invariant and have a unique null in the desired direction. Figure 6 plots the DI and the WNG of the NCMWNG beamformer as a function of frequency, f. As seen, this beamformer designed with $M = 12$ achieves a much higher value of WNG and keeps almost the same DI, which shows the superiority of using more microphones than required to improve robustness.

Finally, we study the differential beamformer that combines both the pressure and difference pressure observations, which is called combined differential beamformer. Figure 7 plots the DF and the WNG of this beamformer with $M = 4$ and

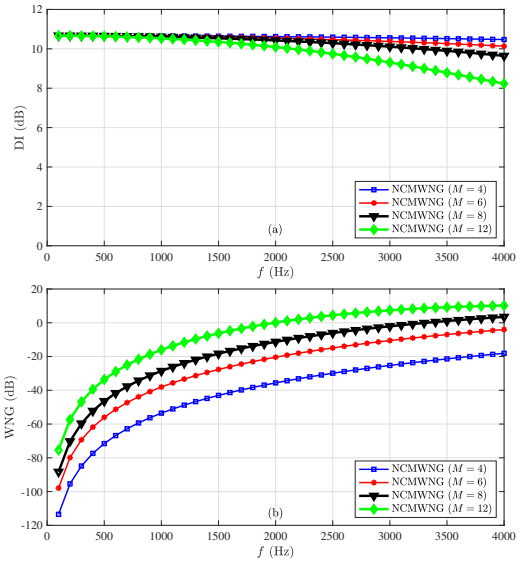

Fig. 6. Performance of the NCMWNG differential beamformer as a function of frequency for different numbers of microphones: (a) DI and (b) WNG. Conditions of simulation: $\delta = 1.0$ cm and $P = 2$.

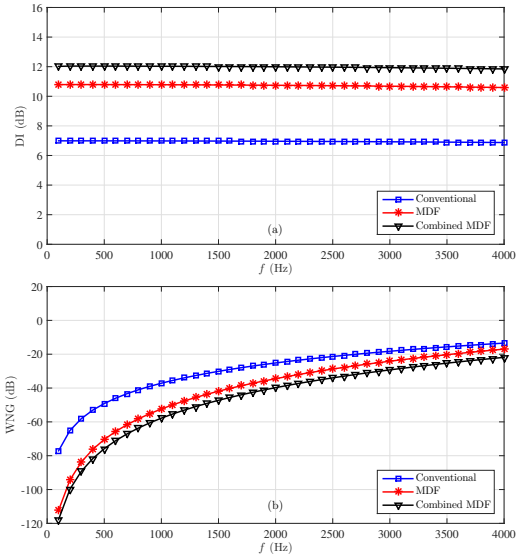

Fig. 7. Performance of the combined MDF differential beamformer as a function of frequency: (a) DI and (b) WNG. Conditions of simulation: $M = 4$ and $\delta = 1.0$ cm.

$P = 2$, as a function of frequency, f. For comparison, we also show plots of the basic forward spatial differential operator and the MDF differential beamformer. It is clearly seen that the combined MDF beamformer can further improve the DI without much sacrifice of the WNG. More importantly, the combined differential beamformer increases the number of degrees of freedom, so it is possible to add more constraints and can be very useful in practice.

B. Performance in Simulated Reverberant Environments

In this subsection, we assess the performance of the proposed differential beamformers in simulated reverberated environments. We consider a room of size $4\ \mathrm{m}\times 5\ \mathrm{m}\times 4\ \mathrm{m}$, where

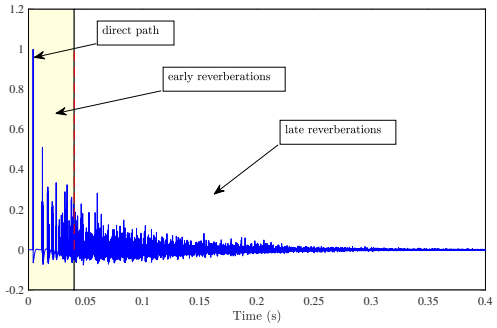

Fig. 8. Acoustic impulse response from the desired source to the first microphone. The reverberation time, T_{60}, is approximately 625 ms, and the sampling rate is 16 kHz.

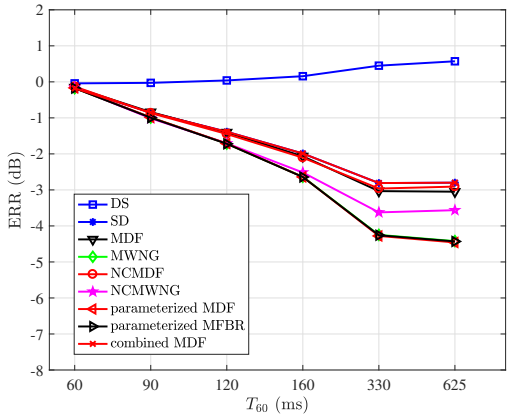

Fig. 9. ERR of the different kinds of differential beamformers for different reverberation conditions.

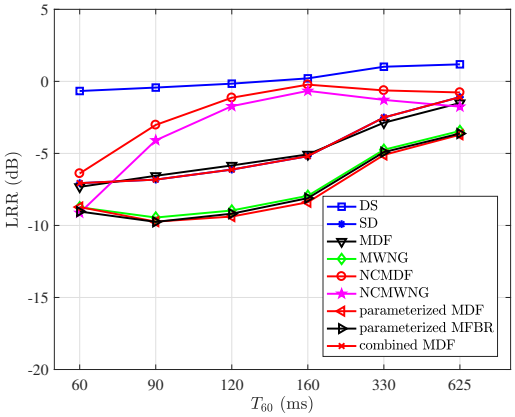

Fig. 10. LRR of the different kinds of differential beamformers for different reverberation conditions.

a linear array of 6 omnidirectional microphones are located, respectively, at $(x, 2.0, 2.0)$, where $x = 1.5335 : 0.023 : 1.64$. A desired point source is placed at $(2.5, 2.0, 1.5)$ (endfire direction) and plays back a speech signal prerecorded from a female speaker in a quiet office room. An interference point source is placed at $(1.58, 3.5, 1.5)$ (90° direction) and also plays back a speech signal prerecorded from a male speaker in a quiet office room. The acoustic channel impulse responses from the source to the microphones are generated with the

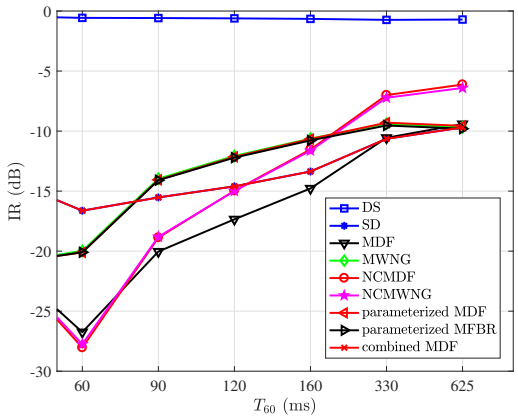

Fig. 11. IR of the different kinds of differential beamformers for different reverberation conditions.

image model method [42]. Then, the microphone observed signals are generated by convolving the source signal with the corresponding impulse responses and adding some noise. In our experiments, we consider two types of noise: white and diffuse. The white noise is a computer generated white Gaussian random process, the diffuse noise has an energy flow of equal probability in all directions [43]. All the signals are sampled at 16 kHz. The overall length of the signal was approximately 10 seconds. The reflection coefficients of all the six walls are are assumed to be identical and frequency independent and they are set to 0, 0.2, 0.4, 0.5, 0.6, 0.8, and 0.9, respectively, and the corresponding reverberation time, T_{60}, are approximately 0 ms, 60 ms, 90 ms, 120 ms, 160 ms, 330 ms, and 625 ms.

For a better study of the beamforming performance, we decompose the observed reverberated desired signal as the direct path signal, which contains the arrival of the direct sound, early reverberation, which contains several strong reflections, and late reverberation, which contains a series of numerous indistinguishable reflections [44]. An example of the acoustic impulse response and the corresponding direct path, early reverberation (the first 40 ms), and late reverberation is shown in Fig. 8. In this case, the microphone observed signal can be written as

$$\begin{aligned} y_m(t) &= g_{s,m}(t) * s(t) + g_{\mathrm{i},m}(t) * i(t) + v_{\mathrm{d},m}(t) + v_{\mathrm{w},m}(t) \\ &= x_m(t) + x_{\mathrm{i},m}(t) + v_{\mathrm{d},m}(t) + v_{\mathrm{w},m}(t) \\ &= x_{\mathrm{d},m}(t) + x_{\mathrm{e},m}(t) + x_{\mathrm{r},m}(t) \\ &\quad + x_{\mathrm{i},m}(t) + v_{\mathrm{d},m}(t) + v_{\mathrm{w},m}(t), \end{aligned} \tag{103}$$

with $m = 1, 2, \ldots, M$, where $*$ stands for linear convolution, $g_{s,m}(t)$ is the acoustic impulse response from the position of the desired source to the mth microphone, $s(t)$ is the desired source signal, $g_{\mathrm{i},m}(t)$ is the acoustic impulse response from the position of the interference source to the mth microphone, $i(t)$ is the interference signal, $x_m(t) = g_{s,m}(t) * s(t)$ is the convolved desired signal at the mth sensor, $x_{\mathrm{i},m}(t) = g_{i,m}(t) * i(t)$ is the convolved interference signal at the mth sensor, and $x_{\mathrm{d},m}(t)$, $x_{\mathrm{e},m}(t)$, and $x_{\mathrm{r},m}(t)$ are the direct path, early reverberation, and late reverberation of the convolved desired signal, respectively. All signals are partitioned into overlapping frames with a frame size of $K = 128$ and an overlapping factor of 75%. A Kaiser window is then applied to each frame and the windowed frame signal is subsequently transformed into the STFT domain using a 128-point FFT. The beamforming filters are then implemented in the STFT domain. Finally, the inverse FFT (with overlap add technique) is used to obtain the time-domain clean speech estimate. Assume the direct path signal, early reverberation, late reverberation, interference, diffuse noise, and white noise after beamforming are expressed as $x_{\mathrm{fd}}(t)$, $x_{\mathrm{fe}}(t)$, $x_{\mathrm{fr}}(t)$, $x_{\mathrm{fi}}(t)$, $v_{\mathrm{fdn}}(t)$, and $v_{\mathrm{fwn}}(t)$, respectively. In our implementation, we set $P = 1$, the matrix $\mathbf{\Gamma}_{\mathrm{d}}$ is regularized by adding to it the small number 10^{-3}. We study the MDF, MWNG, NCMDF, NCMWNG, parameterized MDF ($N = 4$), parameterized MFBR ($N = 4$), and combined MDF differential beamformers. For comparison, we also implemented the widely used delay-and-sum (DS) and superdirective (SD) beamformers. Note that the SD beamformer is a particular case of conventional differential beamformers (when the microphone array size is very small, the superdirective beamformer corresponds to the hypercardioid [16]). We use the fullband time-domain signal to evaluate the overall performance.

We first study the influence of early reverberation on performance, we define the early reverberation reduction (ERR) factor:

$$\mathrm{ERR} = \frac{E\left[x_{\mathrm{fe}}^2(t)\right]}{E\left[x_{\mathrm{e},m}^2(t)\right]}. \tag{104}$$

In our implementation, the first microphone is chosen as the reference, i.e., $m = 1$. Figure 9 plots the ERR for different reverberation conditions. It is clearly seen that the ERR is very small in light reverberation conditions with small reverberation time. As the reverberation time increases, the reduction on early reverberation becomes more significant. The DS beamformer has the smallest ERR in all conditions, which is due to the large beamwidth of the DS beamformer. Note that, as shown in the literature, early reverberation does not affect much the speech quality. So, the amount of reduction in early reverberation may not affect much speech quality.

To evaluate the reduction on late reverberation, we define a late reverberation reduction (LRR) factor:

$$\mathrm{LRR} = \frac{E\left[x_{\mathrm{fr}}^2(t)\right]}{E\left[x_{\mathrm{r},m}^2(t)\right]}. \tag{105}$$

The results for LRR for different reverberation conditions are plotted in Fig. 10. Except the DS beamformer, significant late reverberation reduction is observed in all the studied reverberation conditions. The SD beamformer achieves similar performance as the MDF beamformer. This is understandable as both of them maximize the DF. In comparison, the MWNG, parameterized MDF, and parameterized MFBR beamformers have the lowest LRR, which shows their effectiveness in suppressing late reverberation. The NCMDF and NCMWNG differential beamformers have higher LRR than other beamformers. This is due to the fact that the null-constrained differential beamformers have a very low gain at the null positions, which weakens the effectiveness of the beamformer on other

TABLE I
DSR, DNR, AND WNR OF THE DIFFERENT KINDS OF DIFFERENTIAL BEAMFORMERS.

Method	DSR (dB)	DNR (dB)	WNR (dB)
DS	-0.002	-5.65	-7.72
SD	-0.019	-11.31	7.98
MDF	-0.009	-11.15	8.70
MWNG	0.008	-7.06	-0.29
NCMDF	0.023	-11.21	27.16
NCMWNG	0.032	-8.21	26.67
parameterized MDF	0.006	-7.33	0.96
parameterized MFBR	0.006	-7.54	-0.14
combined MDF	-0.019	-11.31	8.02

directions; but the reverberation is formed by reflections from all directions.

Similarly, we define the interference reduction (IR) factor for the interference signal:

$$\mathrm{IR} = \frac{E\left[x_{\mathrm{fi}}^2(t)\right]}{E\left[x_{\mathrm{i},m}^2(t)\right]}. \tag{106}$$

Figure 11 plots the IR for different reverberation conditions. As seen, all beamformers achieve a good suppression on the interference except for the DS beamformer. The MDF differential beamformer has the overall best performance. The NCMDF and NCMWNG differential beamformers have good performance in light reverberation but slightly worse performance in strong reverberation conditions. The MDF beamformer yields a better performance than the SD beamformer. The underlying reason is that the MDF beamformer has a null at the interference direction. As expected, the IR of all the beamformers increases with the reverberation time, which indicates that higher reverberation will decrease interference suppression ability of beamformers.

We also assess the performance of recovering the desired signal. For that, we define the desired signal reduction (DSR):

$$\mathrm{DSR} = \frac{E\left[x_{\mathrm{fd}}^2(t)\right]}{E\left[x_{\mathrm{d},m}^2(t)\right]}. \tag{107}$$

For the diffuse noise and white noise, we define the diffuse noise reduction (DNR) factor and white noise reduction (WNR) factor, respectively, as

$$\mathrm{DNR} = \frac{E\left[x_{\mathrm{fdn}}^2(t)\right]}{E\left[v_{\mathrm{d},m}^2(t)\right]}, \tag{108}$$

$$\mathrm{WNR} = \frac{E\left[x_{\mathrm{fwn}}^2(t)\right]}{E\left[v_{\mathrm{w},m}^2(t)\right]}. \tag{109}$$

The results are presented in Table I. It is clearly seen that the DSR almost equal to 0 dB for all beamformers while there are subtle differences. That is not surprising since all beamformers are signal independent fixed beamformers in which the distortionless constraint should always be satisfied regardless of the degree of reverberation. The slight differences mainly come from the error of matrix inversion and decomposition calculation. As seen, the SD, MDF, NCMDF, and combined MDF differential beamformers have the lowest DNR while the DS beamformer has the highest DNR but lowest WNR. This is understandable as the DNR and WNR are directly related to the DF and WNG. The SD, MDF, NCMDF, and combined MDF maximize the DF and, as a result, they has the lowest DNR. In comparison, the DS maximizes the WNG. So, it has the lowest WNR. Note that the diffuse noise and white noise are independent with reverberation; so, the DNR and WNG do not change with the reverberation conditions.

In summary, the following conclusions can be drawn. The DS beamformer has the best robustness performance but not effective in dealing with reverberation, interference, and diffuse noise. The SD, MDF, NCMDF, combined MDF, and SD beamformers have the best performance in suppressing diffuse noise but with poor performance in white noise. The MDF differential beamformer has the best performance in suppressing interference and the parameterized MDF, parameterized MFBR, and MWNG differential beamformers have the best performance in suppressing reverberation. In practical systems, it is important for beamformers to have good performance in suppressing reverberation and interference. From this perspective, the MDF, MWNG, parameterized MDF, and parameterized MFBR differential beamformers are superior. They also achieve reasonable performance in diffuse and white noise.

VI. CONCLUSIONS

Differential beamforming is attractive for many acoustic and speech applications with small-size microphone arrays due to its high directional gains and frequency-invariant beampatterns. In this paper, we presented a new theory and some new methods of differential beamforming with uniform linear arrays. We defined a forward spatial difference operator such that any order of the differential observed signal can be represented as a product between a differential matrix and the microphone array observations. The differential beamforming in our method consists of the differential stage, which obtains the differential observed signals, and the filtering stage, which optimizes the beamforming performance. The proposed theory shows the connection between the conventional differential beamforming and null-constrained differential beamforming methods. We deduced new fixed differential beamformers, demonstrated their performance, and presented new insights into the design of differential beamformers. The new differential beamformers can achieve not only tradeoffs between WNG and DF, but also better performance in reverberant acoustic environments. This approach can be further extended to the design of other kinds of differential beamformers.

ACKNOWLEDGEMENTS

The authors thank the associate editor and the anonymous reviewers for their constructive comments and useful suggestions.

REFERENCES

[1] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.

[2] J. Benesty, I. Cohen, and J. Chen, *Array Processing: Kronecker Product Beamforming*, vol. 18. Berlin, Germany: Springer-Verlag, 2019.

[3] B. Rafaely, *Fundamentals of Spherical Array Processing*. Berlin, Germany: Springer-Verlag, 2015.

[4] S. Yan, Y. Ma, and C. Hou, "Optimal array pattern synthesis for broadband arrays," *J. Acoust. Soc. Am.*, vol. 122, no. 5, pp. 2686–2696, 2007.

[5] J. Bitzer and K. U. Simmer, "Superdirective microphone arrays," in *Microphone Arrays*, pp. 19–38, Springer, 2001.

[6] M. Crocco and A. Trucco, "Design of robust superdirective arrays with a tunable tradeoff between directivity and frequency-invariance," *IEEE Trans. Signal Process.*, vol. 59, no. 5, pp. 2169–2181, 2011.

[7] H. F. Olson, "Gradient microphones," *J. Acoust. Soc. Am.*, vol. 17, no. 3, pp. 192–198, 1946.

[8] T. D. Abhayapala and A. Gupta, "Higher order differential-integral microphone arrays," *J. Acoust. Soc. Am.*, vol. 136, pp. 227–233, May 2010.

[9] G. W. Elko, "Superdirectional microphone arrays," in *Acoustic Signal Processing for Telecommunication*, pp. 181–237, Springer, 2000.

[10] E. D. Sena, H. Hacihabiboglu, and Z. Cvetkovic, "On the design and implementation of higher-order differential microphones," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 20, pp. 162–174, Jan. 2012.

[11] G. Huang, J. Chen, and J. Benesty, "Design of planar differential microphone arrays with fractional orders," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 28, no. 1, 2020.

[12] M. Kolundzija, C. Faller, and M. Vetterli, "Spatiotemporal gradient analysis of differential microphone arrays," *Journal Audio Eng. Soc.*, vol. 59, no. 1/2, pp. 20–28, 2011.

[13] G. W. Elko, "Differential microphone arrays," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, pp. 11–65, Springer, 2004.

[14] G. W. Elko, "Microphone array systems for hands-free telecommunication," *Speech Commun.*, vol. 20, no. 3, pp. 229–240, 1996.

[15] M. Buck, "Aspects of first-order differential microphone arrays in the presence of sensor imperfections," *European Trans. Telecomm.*, vol. 13, no. 2, pp. 115–122, 2002.

[16] J. Benesty and J. Chen, *Study and Design of Differential Microphone Arrays*. Berlin, Germany: Springer-Verlag, 2012.

[17] G. Huang, J. Benesty, and J. Chen, "On the design of frequency-invariant beampatterns with uniform circular microphone arrays," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 25, no. 5, pp. 1140–1153, 2017.

[18] G. Huang, J. Benesty, and J. Chen, "Design of robust concentric circular differential microphone arrays," *J. Acoust. Soc. Am.*, vol. 141, no. 5, pp. 3236–3249, 2017.

[19] M. Ihle, "Differential microphone arrays for spectral subtraction," in *Proc. IEEE IWAENC*, pp. 259–262, 2003.

[20] R. M. Derkx and K. Janse, "Theoretical analysis of a first-order azimuth-steerable superdirective microphone array," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 1, pp. 150–162, 2009.

[21] G. W. Elko and A.-T. N. Pong, "A steerable and variable first-order differential microphone array," in *Proc. IEEE ICASSP*, vol. 1, pp. 223–226, IEEE, 1997.

[22] G. W. Elko and J. Meyer, "Microphone arrays," in *Springer Handbook of Speech Processing* (J. Benesty, M. M. Sondhi, and Y. Huang, eds.), ch. 48, pp. 1021–1041, Berlin, Germany: Springer-Verlag, 2008.

[23] H. Teutsch and G. W. Elko, "First-and second-order adaptive differential microphone arrays," in *Proc. IEEE IWAENC*, pp. 35–38, 2001.

[24] X. Wu and H. Chen, "Directivity factors of the first-order steerable differential array with microphone mismatches: deterministic and worst-case analysis," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 24, no. 2, pp. 300–315, 2016.

[25] J. Chen, J. Benesty, and C. Pan, "On the design and implementation of linear differential microphone arrays," *J. Acoust. Soc. Am.*, vol. 136, pp. 3097–3113, Dec. 2014.

[26] G. Huang, J. Chen, and J. Benesty, "On the design of differential beamformers with arbitrary planar microphone array," *J. Acoust. Soc. Am.*, vol. 144, no. 1, pp. 3024–3035, 2018.

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

[28] S. Markovich-Golan and S. Gannot, "Performance analysis of the covariance subtraction method for relative transfer function estimation and comparison to the covariance whitening method," in *Proc. IEEE ICASSP*, pp. 544–548, IEEE, 2015.

[29] N. Ito, S. Araki, M. Delcroix, and T. Nakatani, "Probabilistic spatial dictionary based online adaptive beamforming for meeting recognition in noisy and reverberant environments," in *Proc. IEEE ICASSP*, pp. 681–685, IEEE, 2017.

[30] S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, pp. 2230–2244, Sep. 2002.

[31] N. Ito, S. Araki, and T. Nakatani, "FastFCA: Joint diagonalization based acceleration of audio source separation using a full-rank spatial covariance model," in *eusipco*, pp. 1667–1671, IEEE, 2018.

[32] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Upper Saddle River, NJ: Prentice-Hall, 1993.

[33] P. D. Teal, T. D. Abhayapala, R. A. Kennedy, "Spatial correlation for general distributions of scatterers," in *IEEE signal process. letters*, vol. 9, p. 305-308, Dec. 2002.

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

[35] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 3, pp. 393–398, 1986.

[36] Y. Hu and P. C. Loizou, "A generalized subspace approach for enhancing spech corrupted by colored noise," *IEEE Trans. Speech Audio Processing*, vol. 11, pp. 334–341, July 2003.

[37] J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, pp. 1218–1234, Jul. 2006.

[38] G. Huang, J. Benesty, and J. Chen, "Subspace superdirective beamforming with uniform circular microphone arrays," in *Proc. IEEE IWAENC*, pp. 1–5, 2016.

[39] J. N. Franklin, *Matrix Theory*. Englewood Cliffs, NJ: Prentice-Hall, 1968.

[40] G. Huang, J. Chen, and J. Benesty, "A flexible high directivity beamformer with spherical microphone arrays," *J. Acoust. Soc. Am.*, vol. 143, no. 5, pp. 3024–3035, 2018.

[41] R. N. Marshall and W. R. Harry, "A new microphone providing uniform directivity over an extended frequency range," *J. Acoust. Soc. Am.*, vol. 12, pp. 481–497, 1941.

[42] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating smallroom acoustics," *J. Acoust. Soc. Am.*, vol. 65, p. 943–950, Apr. 1979.

[43] F. Jacobsen, *The diffuse sound field: Statistical considerations concerning the reverberant field in the steady state*. Acoustics Laboratory, Technical University of Denmark, 1979.

[44] D. A. Berkley and J. B. Allen, "Normal listening in typical rooms – the physical and psychophysical correlates of reverberation," in *Acoustical Factors Affecting Hearing Aid Performance* (G.A. Studebaker and I. Hochberg, eds.), Allyn Bacon, Needham Height, 1993.

Gongping Huang (SM'2013) is a postdoctoral research fellow of Electrical Engineering at the Technion – Israel Institute of Technology, Haifa, Israel. He received the Bachelor's degree in Electronics and Information Engineering and the Ph.D. degree in information and communication engineering from the Northwestern Polytechnical University (NPU), Xian, China, in 2012 and 2019, respectively. From 2015 to 2017, he was a visiting researcher at university of Quebec, INRS-EMT, Montreal, Canada.

His research interests include speech enhancement, audio and speech processing, microphone arrays, and graph signal processing. He received the Andrew and Erna Finci Viterbi Post-Doctoral Fellowship's award (2019). He serves as a reviewer for many scientific journals include the IEEE/ACM TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, IEEE SIGNAL PROCESSING LETTER, JOURNAL OF THE ACOUSTICAL SOCIETY OF AMERICA, IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, SPEECH COMMUNICATION, etc.

Jacob Benesty received a Master degree in microwaves from Pierre & Marie Curie University, France, in 1987, and a Ph.D. degree in control and signal processing from Orsay University, France, in April 1991. During his Ph.D. (from Nov. 1989 to Apr. 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecomunications (CNET), Paris, France. From January 1994 to July 1995, he worked at Telecom Paris University on multichannel adaptive filters and acoustic echo cancellation. From October 1995 to May 2003, he was first a Consultant and then a Member of the Technical Staff at Bell Laboratories, Murray Hill, NJ, USA. In May 2003, he joined the University of Quebec, INRS-EMT, in Montreal, Quebec, Canada, as a Professor. He is also a Visiting Professor at the Technion, Haifa, Israel, an Adjunct Professor at Aalborg University, Denmark, and a Guest Professor at Northwestern Polytechnical University, Xi'an, China.

His research interests are in signal processing, acoustic signal processing, and multimedia communications. He is the inventor of many important technologies. In particular, he was the lead researcher at Bell Labs who conceived and designed the world-first real-time hands-free full-duplex stereophonic teleconferencing system. Also, he conceived and designed the world-first PC-based multi-party hands-free full-duplex stereo conferencing system over IP networks.

He is the editor of the book series Springer Topics in Signal Processing. He was the general chair and technical chair of many international conferences and a member of several IEEE technical committees. Four of his journal papers were awarded by the IEEE Signal processing Society and in 2010 he received the Gheorghe Cartianu Award from the Romanian Academy. He has co-authored and co-edited/co-authored numerous books in the area of acoustic signal processing.

Israel Cohen (M'01-SM'03-F'15) is a Professor of electrical engineering at the Technion – Israel Institute of Technology, Haifa, Israel. He is also a Visiting Professor at Northwestern Polytechnical University, Xi'an, China. He received the B.Sc. (Summa Cum Laude), M.Sc. and Ph.D. degrees in electrical engineering from the Technion – Israel Institute of Technology, in 1990, 1993 and 1998, respectively.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT, USA. In 2001 he joined the Electrical Engineering Department of the Technion. He is a coeditor of the Multichannel Speech Processing Section of the *Springer Handbook of Speech Processing* (Springer, 2008), and a coauthor of *Fundamentals of Signal Enhancement and Array Signal Processing* (Wiley-IEEE Press, 2018). His research interests are array processing, statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification and adaptive filtering.

Dr. Cohen was awarded the Norman Seiden Prize for Academic Excellence (2017), the SPS Signal Processing Letters Best Paper Award (2014), the Alexander Goldberg Prize for Excellence in Research (2010), and the Muriel and David Jacknow Award for Excellence in Teaching (2009). He serves as Associate Member of the IEEE Audio and Acoustic Signal Processing Technical Committee, and as Distinguished Lecturer of the IEEE Signal Processing Society. He served as Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS, and as Member of the IEEE Audio and Acoustic Signal Processing Technical Committee and the IEEE Speech and Language Processing Technical Committee.

Jingdong Chen (M'99–SM'09) received the Ph.D. degree in pattern recognition and intelligence control from the Chinese Academy of Sciences in 1998.

From 1998 to 1999, he was with ATR Interpreting Telecommunications Research Laboratories, Kyoto, Japan, where he conducted research on speech synthesis, speech analysis, as well as objective measurements for evaluating speech synthesis. He then joined the Griffith University, Brisbane, Australia, where he engaged in research on robust speech recognition and signal processing. From 2000 to 2001, he worked at ATR Spoken Language Translation Research Laboratories on robust speech recognition and speech enhancement. From 2001 to 2009, he was a Member of Technical Staff at Bell Laboratories, Murray Hill, New Jersey, working on acoustic signal processing for telecommunications. He subsequently joined WeVoice Inc. in New Jersey, serving as the Chief Scientist. He is currently a professor at the Northwestern Polytechnical University in Xi'an, China. His research interests include array signal processing, adaptive signal processing, speech enhancement, adaptive noise/echo control, signal separation, speech communication, and artificial intelligence.

Dr. Chen served as an Associate Editor of the IEEE Transactions on Audio, Speech, and Language Processing from 2008 to 2014 and as a technical committee (TC) member of the IEEE Signal Processing Society (SPS) TC on Audio and Electroacoustics from 2007 to 2009. He is currently a member of the IEEE SPS TC on Audio and Acoustic Signal Processing, and a member of the editorial advisory board of the Open Signal Processing Journal. He was the General Co-Chair of ACM WUWNET 2018 and IWAENC 2016, the Technical Program Chair of IEEE TENCON 2013, a Technical Program Co-Chair of IEEE WASPAA 2009, IEEE ChinaSIP 2014, IEEE ICSPCC 2014, and IEEE ICSPCC 2015, and helped organize many other conferences. He co-authored 12 monograph books including *Array Processing–Kronecker Product Beamforming*, (Springer, 2019), *Fundamentals of Signal Enhancement and Array Signal Processing*, (Wiley, 2018), *Fundamentals of Differential Beamforming*, (Springer, 2016), *Design of Circular Differential Microphone Arrays* (Springer, 2015), *Noise Reduction in Speech Processing* (Springer, 2009), *Microphone Array Signal Processing* (Springer, 2008), and *Acoustic MIMO Signal Processing* (Springer, 2006), etc.

Dr. Chen received the 2008 Best Paper Award from the IEEE Signal Processing Society (with Benesty, Huang, and Doclo), the Best Paper Award from the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics in 2011 (with Benesty), the Bell Labs Role Model Teamwork Award twice, respectively, in 2009 and 2007, the NASA Tech Brief Award twice, respectively, in 2010 and 2009, and the Young Author Best Paper Award from the 5th National Conference on Man-Machine Speech Communications in 1998. He is a co-author of a paper for which C. Pan received the IEEE R10 (Asia-Pacific Region) Distinguished Student Paper Award (First Prize) in 2016. He was also a recipient of the Japan Trust International Research Grant from the Japan Key Technology Center in 1998 and the "Distinguished Young Scientists Fund" from the National Natural Science Foundation of China (NSFC) in 2014.