A multistage minimum variance distortionless response beamformer for noise reduction

Chao Pan and Jingdong Chen[a)]
Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, 127 Youyi West Road, Xi'an, Shaanxi 710072, China

Jacob Benesty
Institut National de la Recherche Scientifique-Énergie Matériaux Télécommunications, University of Quebec, 800 de la Gauchetiere Ouest, Suite 6900, Montreal, Quebec H5A 1K6, Canada

(Received 6 December 2014; accepted 12 February 2015)

This paper develops a multistage approach to the implementation of the minimum variance distortionless response (MVDR) beamformer. It first divides the microphone array of M sensors into $M/2$ subarrays with each subarray having only two microphones, and a two-channel MVDR beamformer is performed with each subarray. The $M/2$ subarrays' outputs are then treated as the inputs of $M/4$ subarrays of two channels in the next stage. Similarly, a two-channel MVDR beamformer is performed with each subarray in the second stage. This process is repeated till the last stage that has only a single output. This multistage MVDR beamformer has the following properties: (1) Its performance is identical to that of the conventional MVDR beamformer in spatially uncorrelated noise; (2) it is much more robust than the conventional MVDR beamformer in diffuse noise, i.e., it has a significantly higher white noise gain as compared to the traditional MVDR beamformer; and (3) its complexity is an order of magnitude smaller than that of the traditional MVDR beamformer. This basic principle can also be easily generalized to the case where every subarray has more than two microphones. © *2015 Acoustical Society of America.*
[http://dx.doi.org/10.1121/1.4913459]

[MRB] Pages: 1377–1388

I. INTRODUCTION

The minimum variance distortionless response (MVDR) beamformer, originally developed by Capon (1969), has been widely studied in the context of noise reduction and speech enhancement with microphone arrays to extract the speech signal of interest and reduce the unwanted noise. The robust and efficient implementation of this beamformer, however, is not a trivial task because the correlation matrices of the noise and noisy signals that need to be inverted are usually ill-conditioned. If not handled properly, this ill-conditioning issue can cause numerical instabilities of the MVDR beamformer, which may lead to significant white noise amplification. A great deal of effort has been devoted to circumventing this issue, which can be categorized into two classes: Diagonal loading (DL) and robust adaptation structure.

The basic idea of DL is to add a small positive constant to the diagonal elements of the noise or noisy correlation matrix, making it better conditioned (Li *et al.,* 2003; Carlson, 1988). This is equivalent to adding some amount of white Gaussian noise to the array observations. The critical issue with this technique is the choice of the proper value of the loading parameter. On the one hand, the ill-conditioning issue would remain if the loading constant is too small, and on the other hand, the array directivity may degrade significantly if the loading constant is too large. In the context of noise reduction with wideband speech signals, diagonal loading may cause the array to have a different response from one frequency to another; this will be perceived as speech distortion. Another equivalent method to diagonal loading is the so-called dominant mode rejection (DMR) method (Cox and Pitre, 1998; Kogon, 2004). The basic idea is to determine the dominant eigenvalues of the noise or noisy correlation matrices. Then either the pseudoinverse is used to replace the direct inverse (Pan *et al.,* 2014) or the beamformer can be reformulated to reject noise at the space spanned by the eigenvectors corresponding to the dominant eigenvalues. Similar to the DL method, the DMR technique also suffers from directivity degradation.

The robustness of the MVDR beamformer may also be slightly improved through the use of a different, robust adaptation structure. It is well known that the generalized sidelobe canceller (GSC) converts the constrained optimization problem in the linear constrained minimum variance (LCMV) beamformer into an unconstrained one (Griffiths and Jim, 1982; Gannot *et al.,* 2001). While they are theoretically identical (Breed and Strauss, 2004), the GSC may be slightly more robust than the LCMV in implementation. Because it is a particular case of the LCMV (Frost, 1972), the MVDR beamformer can be implemented with the GSC structure. Another way to implement the MVDR beamformer is through the iterative method developed in (Pados and Karystinos, 2001), where the MVDR filter is iteratively updated from a matched filter combined with a filter that is orthogonal to the matched filter. Similar to the GSC, this iterative method avoids the direct computation of matrix

[a)]Author to whom correspondence should be addressed. Electronic mail: jingdongchen@ieee.org

inversion, and it can converge to the MVDR beamformer in theory. However, this method was found sensitive to the estimation error of the correlation matrix, particularly in low frequencies, where the correlation matrix is generally ill-conditioned. Moreover, the computational complexity of this algorithm is high.

In this paper, we develop a multistage MVDR beamformer as illustrated in Fig. 1. Briefly, we first divide the microphone array of M sensors into $M/2$ subarrays where every subarray has only two microphones. A two-channel MVDR beamformer is performed with each subarray. The $M/2$ subarrays' outputs are then treated as the inputs of the second stage with $M/4$ subarrays of two channels. Similarly, a two-channel MVDR beamformer is performed with each subarray in the second stage. This process is repeated till the last stage that has two inputs and only a single output. We will present the theoretical analysis of this multistage MVDR beamformer and show that this approach has many appealing properties including: (1) Its performance is identical to that of the conventional MVDR beamformer in spatially uncorrelated noise; (2) it has significantly higher white noise gain in diffuse noise, meaning that it is much more robust than the conventional MVDR beamformer though its array gain is slightly smaller; and (3) its complexity is an order of magnitude smaller than that of the conventional MVDR beamformer. This basic principle can also be easily generalized to the case where every subarray has more than two microphones. As the number of sensors in each subarray increases, the performance behavior of the multistage MVDR beamformer gets closer to the conventional one; but the robustness decreases while the complexity increases.

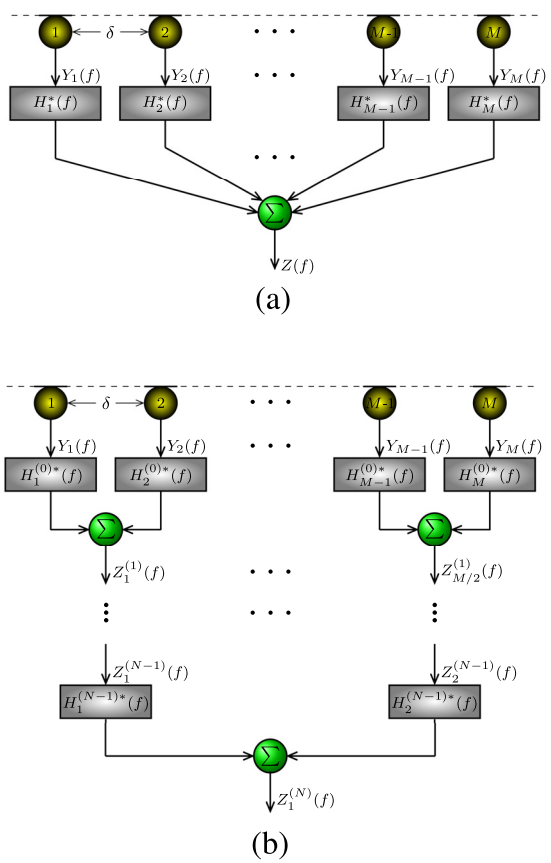

FIG. 1. (Color online) Schematic diagrams of the conventional and multistage MVDR beamformers.

The rest of this paper is organized as follows. In Sec. II, the general signal model in the frequency domain is presented. Section III discusses the conventional MVDR beamformer. In Sec. IV, the multistage MVDR beamformer is derived, and the corresponding complexity is analyzed. Section V presents some important performance metrics for the evaluation of the conventional and multistage MVDR beamformers. Then Sec. VI studies the performance of the multistage MVDR beamformer in different scenarios and compares it to that of the conventional MVDR beamformer. Finally, some conclusions are provided in Sec. VII.

II. SIGNAL MODEL

We consider the well-accepted room acoustics signal model in which an M-element microphone array captures a convolved source signal in some noise field. The received signals, at the time index t, are expressed as (Benesty *et al.*, 2008; Brandstein and Ward, 2001)

$$\begin{aligned} y_m(t) &= g_m(t) * s(t) + v_m(t) \\ &= x_m(t) + v_m(t), \quad m = 1, 2, \ldots, M, \end{aligned} \tag{1}$$

where $g_m(t)$ is the impulse response from the unknown speech source $s(t)$ to the mth microphone, $*$ stands for linear convolution, and $v_m(t)$ is the additive noise at microphone m. We assume that the signals $x_m(t) = g_m(t) * s(t)$ and $v_m(t)$ are uncorrelated and zero mean. By definition, the signals $x_m(t)$, $m = 1, 2, \ldots, M$ are coherent. The noise components, $v_m(t)$, $m = 1, 2, \ldots, M$, however, are assumed not to be completely coherent. All previous signals are considered to be real and broadband.

In this paper, our desired signal is designated by the clean (but convolved) speech signal received at microphone 1, namely, $x_1(t)$. It should be noted that any other microphone could be considered as the reference. Our problem then may be stated as follows Benesty *et al.* (2008): Given the M observation signals $y_m(t)$, $m = 1, 2, \ldots, M$, our aim is to extract $x_1(t)$. This extraction should be done in such a way that the signal of interest is not much distorted (ideally undistorted) while the noise terms, $v_m(t)$, $m = 1, 2, \ldots, M$, are minimized at the array output.

Expression (1) can be written in the frequency domain, at the frequency index f, as

$$\begin{aligned} Y_m(f) &= G_m(f)S(f) + V_m(f) \\ &= X_m(f) + V_m(f), \quad m = 1, 2, \ldots, M, \end{aligned} \tag{2}$$

where $Y_m(f)$, $G_m(f)$, $S(f)$, $X_m(f) = G_m(f)S(f)$, and $V_m(f)$ are the frequency-domain representations of $y_m(t)$, $g_m(t)$, $s(t)$, $x_m(t)$, and $v_m(t)$, respectively.

To derive the MVDR filter, it is more convenient to write the M frequency-domain microphone signals in a vector form as

$$\begin{aligned}\mathbf{y}(f) &= \mathbf{g}(f)S(f) + \mathbf{v}(f)\\ &= \mathbf{x}(f) + \mathbf{v}(f)\\ &= \mathbf{d}(f)X_1(f) + \mathbf{v}(f),\end{aligned} \tag{3}$$

where

$$\begin{aligned}\mathbf{y}(f) &\triangleq \left[Y_1(f)\ \ Y_2(f)\ \ \cdots\ \ Y_M(f)\right]^T,\\ \mathbf{g}(f) &\triangleq \left[G_1(f)\ \ G_2(f)\ \ \cdots\ \ G_M(f)\right]^T,\\ \mathbf{x}(f) &\triangleq \left[X_1(f)\ \ X_2(f)\ \ \cdots\ \ X_M(f)\right]^T\\ &= S(f)\left[G_1(f)\ \ G_2(f)\ \ \cdots\ \ G_M(f)\right]^T\\ &= S(f)\mathbf{g}(f),\\ \mathbf{v}(f) &\triangleq \left[V_1(f)\ \ V_2(f)\ \ \cdots\ \ V_M(f)\right]^T,\\ \mathbf{d}(f) &\triangleq \frac{\mathbf{g}(f)}{G_1(f)},\end{aligned}$$

and superscript T denotes transpose of a vector or a matrix. The vector $\mathbf{d}(f)$ is called the steering vector or direction vector because it determines the direction of the (desired) signal $X_1(f)$. For a uniform linear array and when the source is in the farfield, in an anechoic environment, and arrives at the array with an incidence angle of θ_{d}, the steering vector $\mathbf{d}(f)$ can be written as

$$\mathbf{d}(f) = \left[1\ \ e^{-\jmath 2\pi f\tau_0\cos\theta_{\mathrm{d}}}\ \ \cdots\ \ e^{-\jmath(M-1)2\pi f\tau_0\cos\theta_{\mathrm{d}}}\right]^T, \tag{4}$$

where $\tau_0 = \delta/c$ is the delay between two successive sensors at the angle $\theta_{\mathrm{d}} = 0^\circ$, δ is the sensor spacing, and c is the sound velocity in air.

From Eq. (3), we easily deduce the correlation matrix of $\mathbf{y}(f)$, which is

$$\begin{aligned}\mathbf{\Phi}_{\mathbf{y}}(f) &\triangleq E[\mathbf{y}(f)\mathbf{y}^H(f)]\\ &= \mathbf{\Phi}_{\mathbf{x}}(f) + \mathbf{\Phi}_{\mathbf{v}}(f)\\ &= \phi_{X_1}(f)\mathbf{d}(f)\mathbf{d}^H(f) + \mathbf{\Phi}_{\mathbf{v}}(f),\end{aligned} \tag{5}$$

where $E[\cdot]$ is the mathematical expectation, the superscript H denotes the conjugate-transpose operator, $\phi_{X_m}(f) \triangleq E[|X_m(f)|^2]$ is the variance of $X_m(f)$, $m = 1, 2,\ldots, M$, and $\mathbf{\Phi}_{\mathbf{x}}(f) \triangleq E[\mathbf{x}(f)\mathbf{x}^H(f)]$ and $\mathbf{\Phi}_{\mathbf{v}}(f) = E[\mathbf{v}(f)\mathbf{v}^H(f)]$ are the correlation matrices of $\mathbf{x}(f)$ and $\mathbf{v}(f)$, respectively. The $M \times M$ matrix $\mathbf{\Phi}_{\mathbf{y}}(f)$ is the sum of two other matrices: The signal correlation matrix with rank of 1 and the noise correlation matrix, which is assumed to be full rank.

III. ARRAY MODEL AND THE CONVENTIONAL MVDR BEAMFORMER

The conventional frequency-domain beamforming is performed by applying a complex weight to the output of each sensor and then summing the results together, i.e.,

$$\begin{aligned}Z(f) &= \sum_{m=1}^{M} H_m^*(f)Y_m(f)\\ &= \mathbf{h}^H(f)\mathbf{y}(f)\\ &= X_1(f)\mathbf{h}^H(f)\mathbf{d}(f) + \mathbf{h}^H(f)\mathbf{v}(f)\\ &= X_{\mathrm{fd}}(f) + V_{\mathrm{rn}}(f),\end{aligned} \tag{6}$$

where the superscript * is the complex-conjugation operator, $Z(f)$ is supposed to be the estimate of $X_1(f)$,

$$\mathbf{h}(f) \triangleq \left[H_1(f)\ \ H_2(f)\ \ \cdots\ \ H_M(f)\right]^T$$

is a filter of length M containing all the complex gains applied to the microphone outputs, $X_{\mathrm{fd}}(f) = X_1(f)\mathbf{h}^H(f)\mathbf{d}(f)$ is the filtered desired signal, and $V_{\mathrm{rn}}(f) = \mathbf{h}^H(f)\mathbf{v}(f)$ is the residual noise.

The two terms on the right-hand side of Eq. (6) are incoherent. Hence the variance of $Z(f)$ is also the sum of two variances

$$\begin{aligned}\phi_Z(f) &= \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{y}}(f)\mathbf{h}(f)\\ &= \phi_{X_{\mathrm{fd}}}(f) + \phi_{V_{\mathrm{rn}}}(f),\end{aligned} \tag{7}$$

where $\phi_{X_{\mathrm{fd}}}(f) = \phi_{X_1}(f)|\mathbf{h}^H(f)\mathbf{d}(f)|^2$ and $\phi_{V_{\mathrm{rn}}}(f) = \mathbf{h}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}(f)$.

Minimizing the variance of the array output or the variance of the residual noise with the constraint that $\mathbf{h}^H(f)\mathbf{d}(f) = 1$ leads to the conventional MVDR beamformer (Capon, 1969; Lacoss, 1971),

$$\begin{aligned}\mathbf{h}_{\mathrm{CMVDR}}(f) &= \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f)}{\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f)}\\ &= \frac{\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{d}(f)}{\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{y}}^{-1}(f)\mathbf{d}(f)}.\end{aligned} \tag{8}$$

In practice, $\mathbf{\Phi}_{\mathbf{x}}(f)$ is rarely a rank-one matrix; as a result, estimating $\mathbf{d}(f)$ is very challenging, which would cause significant performance degradation of the filter. In Benesty *et al.* (2008), the MVDR filter is rewritten as

$$\mathbf{h}_{\mathrm{CMVDR}}(f) = \frac{\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f) - \mathbf{I}_M}{\mathrm{tr}\left[\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{\Phi}_{\mathbf{y}}(f)\right] - M}\mathbf{i}_M, \tag{9}$$

where $\mathrm{tr}[\cdot]$ is the trace of a square matrix, $\mathbf{I}_M$ is the identity matrix of size $M \times M$, and $\mathbf{i}_M$ is the first column of $\mathbf{I}_M$. The two versions of the MVDR filter in Eqs. (8) and (9) are theoretically identical, yet the MVDR filter in Eq. (9) is more practical than that in Eq. (8) as it can be evaluated from the statistics of the noise and noisy signal vectors $\mathbf{v}(f)$ and $\mathbf{y}(f)$ that can be observable. We see from Eq. (9) that the $M \times M$ correlation matrix $\mathbf{\Phi}_{\mathbf{v}}(f)$ needs to be inverted, and this inversion may cause some problems to the estimation of the desired signal if this matrix is ill-conditioned. Furthermore, the inversion of large matrices is not always possible in real-time systems.

From now on, we will drop the subscript CMVDR from all variables to simplify the notation and make the presentation concise. Substituting Eq. (8) into Eqs. (6) and (7), we find that

$$Z(f) = X_1(f) + V_{\mathrm{rn}}(f) \tag{10}$$

and

$$\phi_Z(f) = \phi_{X_1}(f) + \phi_{V_{\mathrm{rn}}}(f), \tag{11}$$

where

$$V_{\mathrm{rn}}(f) = \frac{\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{v}(f)}{\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f)},$$
$$\phi_{V_{\mathrm{rn}}}(f) = \frac{1}{\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f)},$$

with $\phi_{V_m}(f) = E\,[|V_m(f)|^2]$ being the variance of $V_m(f)$, $m = 1, 2, \ldots, M$. It can be shown that the signal estimate $Z(f)$ from Eq. (10) is less noisy than $Y_1(f) = X_1(f) + V_1(f)$. Indeed, according to the Cauchy–Schwarz inequality, we have

$$|\mathbf{d}^H(f)\mathbf{i}_M|^2 \leq [\mathbf{i}_M^T\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{i}_M] \times [\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f)].$$

Because of the facts that $\mathbf{i}_M^T\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{i}_M = \phi_{V_1}(f)$ and $\mathbf{d}^H(f)\mathbf{i}_M = 1$, we deduce that $\phi_{V_{\mathrm{rn}}}(f) \leq \phi_{V_1}(f)$.

Interestingly, $Z(f)$ can be seen as a new observation signal from microphone 1, which is less noisy than the original observation. This observation is used in Sec. IV to derive a multistage beamforming algorithm with matrix dimensions much smaller than and independent of M.

IV. MULTISTAGE MVDR BEAMFORMER

A. Derivation

In the previous section, we showed how to estimate $X_1(f)$ with no distortion from all the observations, $Y_m(f)$, $m = 1, 2, \ldots, M$, to form $Z(f)$, which can also be seen as a less noisy observation than $Y_1(f)$. However, in the implementation of the conventional MVDR beamformer, a matrix of size $M \times M$ needs to be inverted at every frequency. It is well known that this matrix may not be well conditioned, and, as a result, the larger is M the less reliable is the algorithm for noise reduction. Furthermore, such an algorithm is computationally expensive to implement in the context of noise reduction as many matrices of size $M \times M$ have to be inverted every few milliseconds. In this section, we propose a new approach where only vectors of length 2 and matrices of size 2×2 are handled at N successive stages with $N = \log M/\log 2$. In total, the proposed algorithm computes $M - 1 = 2^N - 1$ two-channel MVDR filters. In the first stage, 2^{N-1} two-channel MVDR filters are evaluated, in the second stage, 2^{N-2}, and so on, till the last stage with 2^0.

Let us first form pairs of microphone signals, from which we form the two-dimensional observation vectors

$$\mathbf{z}_i^{(0)}(f) = \begin{bmatrix} Y_{2i-1}(f) & Y_{2i}(f) \end{bmatrix}^T = \mathbf{d}_i(f)X_i(f) + \mathbf{v}_i^{(0)}(f), \quad i = 1, 2, \ldots, M/2, \tag{12}$$

where

$$\mathbf{v}_i^{(0)}(f) = [V_{2i-1}(f) \;\; V_{2i}(f)]^T$$

and

$$\mathbf{d}_i(f) = \begin{bmatrix} \dfrac{G_{2i-1}(f)}{G_i(f)} & \dfrac{G_{2i}(f)}{G_i(f)} \end{bmatrix}^T.$$

At stage 1, the output signals of the $M/2$ beamformers are

$$\begin{aligned} Z_i^{(1)}(f) &= \mathbf{h}_i^{(0)H}(f)\mathbf{z}_i^{(0)}(f) \\ &= X_i(f) + \mathbf{h}_i^{(0)H}(f)\mathbf{v}_i^{(0)}(f) \\ &= X_i(f) + V_i^{(1)}(f), \quad i = 1, 2, \ldots, M/2, \end{aligned} \tag{13}$$

where $Z_i^{(1)}(f)$ are the estimates of $X_i(f)$, $i = 1, 2, \ldots, M/2$ at stage 1 (that will become the new observations for the next stage), $\mathbf{h}_i^{(0)}(f) = \left[H_{2i-1}^{(0)} \; H_{2i}^{(0)}\right]^T$ are the MVDR filters (of length 2) corresponding to the estimates $Z_i^{(1)}(f)$, i.e.,

$$\begin{aligned} \mathbf{h}_i^{(0)}(f) &= \frac{\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}^{-1}(f)\mathbf{d}_i(f)}{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}^{-1}(f)\mathbf{d}_i(f)} \\ &= \frac{\mathbf{\Phi}_{\mathbf{z}_i^{(0)}}^{-1}(f)\mathbf{d}_i(f)}{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{z}_i^{(0)}}^{-1}(f)\mathbf{d}_i(f)}, \end{aligned} \tag{14}$$

or, alternatively,

$$\mathbf{h}_i^{(0)}(f) = \frac{G_i(f)}{G_{2i-1}(f)} \times \frac{\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}^{-1}(f)\mathbf{\Phi}_{\mathbf{z}_i^{(0)}}(f) - \mathbf{I}}{\mathrm{tr}\left[\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}^{-1}(f)\mathbf{\Phi}_{\mathbf{z}_i^{(0)}}(f)\right] - 2}\mathbf{i}, \tag{15}$$

$\mathbf{I}$ is the 2×2 identity matrix, $\mathbf{i}$ is the first column of $\mathbf{I}$,

$$\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}(f) = E[\mathbf{v}_i^{(0)}(f)\mathbf{v}_i^{(0)H}(f)], \tag{16}$$

$$\mathbf{\Phi}_{\mathbf{z}_i^{(0)}}(f) = \phi_{X_i}(f)\mathbf{d}_i(f)\mathbf{d}_i^H(f) + \mathbf{\Phi}_{\mathbf{v}_i^{(0)}}(f), \tag{17}$$

and

$$\begin{aligned} V_i^{(1)}(f) &= \mathbf{h}_i^{(0)H}(f)\mathbf{v}_i^{(0)}(f) \\ &= \frac{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}^{-1}(f)\mathbf{v}_i^{(0)}(f)}{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}^{-1}(f)\mathbf{d}_i(f)} \end{aligned} \tag{18}$$

being the residual noise at stage 1 for the new observation $Z_i^{(1)}(f)$.

The transfer function ratio in Eq. (15) can be rewritten as

$$\begin{aligned} \frac{G_i(f)}{G_{2i-1}(f)} &= \frac{E\left[X_{2i-1}^*(f)X_i(f)\right]}{E\left[|X_{2i-1}(f)|^2\right]} \\ &= \frac{E\left[Y_{2i-1}^*(f)Y_i(f)\right] - E\left[V_{2i-1}^*(f)V_i(f)\right]}{E\left[|Y_{2i-1}(f)|^2\right] - E\left[|V_{2i-1}(f)|^2\right]}, \end{aligned} \tag{19}$$

which can be estimated from the statistics of the noisy and noise signals in practice.

The variance of $Z_i^{(1)}(f)$ is

$$\phi_{Z_i^{(1)}}(f) = \phi_{X_i}(f) + \phi_{V_i^{(1)}}(f), \tag{20}$$

where

$$\phi_{V_i^{(1)}}(f) = \frac{1}{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{v}_i^{(0)}}^{-1}(f)\mathbf{d}_i(f)}$$

and

$$\phi_{V_i^{(1)}}(f) \leq \phi_{V_i^{(0)}}(f).$$

At stage 2, we form the new two-dimensional observation vectors as

$$\begin{aligned}\mathbf{z}_i^{(1)}(f) &= \left[Z_{2i-1}^{(1)}(f) \;\; Z_{2i}^{(1)}(f)\right]^T \\ &= \mathbf{d}_i(f)X_i(f) + \mathbf{v}_i^{(1)}(f), \quad i = 1,2,\ldots,M/4, \end{aligned} \tag{21}$$

where

$$\mathbf{v}_i^{(1)}(f) = \left[V_{2i-1}^{(1)}(f) \;\; V_{2i}^{(1)}(f)\right]^T.$$

We then proceed as before to estimate $Z_i^{(2)}(f)$.

Therefore the general procedure at any stage, $n = 1, 2,\ldots, N$, is

$$\begin{aligned} Z_i^{(n)}(f) &= \mathbf{h}_i^{(n-1)H}(f)\mathbf{z}_i^{(n-1)}(f) \\ &= X_i(f) + \mathbf{h}_i^{(n-1)H}(f)\mathbf{v}_i^{(n-1)}(f) \\ &= X_i(f) + V_i^{(n)}(f), \quad i = 1,2,\ldots,2^{N-n}, \end{aligned} \tag{22}$$

where

$$\begin{aligned}\mathbf{z}_i^{(n-1)}(f) &= \left[Z_{2i-1}^{(n-1)}(f) \;\; Z_{2i}^{(n-1)}(f)\right]^T \\ &= \mathbf{d}_i(f)X_i(f) + \mathbf{v}_i^{(n-1)}(f), \end{aligned} \tag{23}$$

$$\mathbf{v}_i^{(n-1)}(f) = \left[V_{2i-1}^{(n-1)}(f) \;\; V_{2i}^{(n-1)}(f)\right]^T, \tag{24}$$

and

$$\mathbf{h}_i^{(n-1)}(f) = \frac{G_i(f)}{G_{2i-1}(f)} \times \frac{\mathbf{\Phi}_{\mathbf{v}_i^{(n-1)}}^{-1}(f)\mathbf{\Phi}_{\mathbf{z}_i^{(n-1)}}(f) - \mathbf{I}}{\mathrm{tr}\left[\mathbf{\Phi}_{\mathbf{v}_i^{(n-1)}}^{-1}(f)\mathbf{\Phi}_{\mathbf{z}_i^{(n-1)}}(f)\right] - 2}\mathbf{i}, \tag{25}$$

$$\begin{aligned} V_i^{(n)}(f) &= \mathbf{h}_i^{(n-1)H}(f)\mathbf{v}_i^{(n-1)}(f) \\ &= \frac{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{v}_i^{(n-1)}}^{-1}(f)\mathbf{v}_i^{(n-1)}(f)}{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{v}_i^{(n-1)}}^{-1}(f)\mathbf{d}_i(f)}, \end{aligned} \tag{26}$$

$$\mathbf{\Phi}_{\mathbf{v}_i^{(n-1)}}(f) = E[\mathbf{v}_i^{(n-1)}(f)\mathbf{v}_i^{(n-1)H}(f)], \tag{27}$$

$$\mathbf{\Phi}_{\mathbf{z}_i^{(n-1)}}(f) = \phi_{X_i}(f)\mathbf{d}_i(f)\mathbf{d}_i^H(f) + \mathbf{\Phi}_{\mathbf{v}_i^{(n-1)}}(f). \tag{28}$$

B. Implementation

For clarity, we stack the noise signals at the nth stage into a vector

$$\mathbf{v}^{(n)}(f) = \left[V_1^{(n)}(f) \;\; V_2^{(n)}(f) \;\; \cdots \;\; V_{2^{N-n}}^{(n)}(f)\right]^T, \tag{29}$$

and define the corresponding correlation matrix

$$\mathbf{\Phi}_{\mathbf{v}^{(n)}}(f) \triangleq E[\mathbf{v}^{(n)}(f)\mathbf{v}^{(n)H}(f)]. \tag{30}$$

After the derivation of the filters at the nth stage, the correlation matrix of the noise signals at the $(n+1)$th stage can be expressed as

$$\mathbf{\Phi}_{\mathbf{v}^{(n+1)}}(f) = \mathbf{H}^{(n)}(f)\mathbf{\Phi}_{\mathbf{v}^{(n)}}(f)\mathbf{H}^{(n)H}(f), \tag{31}$$

where

$$\mathbf{H}^{(n)}(f) = \begin{bmatrix} \mathbf{h}_1^{(n)H} & \mathbf{0}_{1\times 2} & \cdots & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \mathbf{h}_2^{(n)H} & \cdots & \mathbf{0}_{1\times 2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & \cdots & \mathbf{h}_{2^{N-n-1}}^{(n)H} \end{bmatrix}, \tag{32}$$

which consists of the filters at the nth stage and is a matrix of size $2^{N-n-1} \times 2^{N-n}$.

In Table I, we illustrate the multistage algorithm. From this table, the final array output can be written as

$$Z^{(N)}(f) = \mathbf{h}_{\mathrm{MMVDR}}^H(f)\mathbf{y}(f), \tag{33}$$

where

$$\mathbf{h}_{\mathrm{MMVDR}}(f) = \mathbf{H}^{(0)H}(f)\mathbf{H}^{(1)H}(f)\cdots\mathbf{H}^{(N-1)H}(f), \tag{34}$$

which is the multistage MVDR beamformer. The corresponding residual noise variance can be expressed as

$$\phi_{V_{\mathrm{m}}}(f) = \mathbf{h}_{\mathrm{MMVDR}}^H(f)\mathbf{\Phi}_{\mathbf{v}}(f)\mathbf{h}_{\mathrm{MMVDR}}(f). \tag{35}$$

TABLE I. Multistage MVDR beamformer.

Multistage MVDR beamformer

Inputs:
$\mathbf{\Phi}_{\mathbf{v}^{(0)}}(f) = \mathbf{\Phi}_{\mathbf{v}}(f)$, which is the correlation matrix of the noise.
$\mathbf{z}^{(0)}(f) = \mathbf{y}(f)$, which is the observation signal vector.

Filtering:
for $n = 0, 1,\ldots, N-1$
  for $i = 1, 2,\ldots, 2^{N-n-1}$
    $\mathbf{\Phi}_{\mathbf{v}_i^{(n)}}(f) = [\mathbf{\Phi}_{\mathbf{v}^{(n)}}(f)]_{2i-1:2i,2i-1:2i}$
    $\mathbf{h}_i^{(n)}(f) = \dfrac{\mathbf{\Phi}_{\mathbf{v}_i^{(n)}}^{-1}(f)\mathbf{d}_i(f)}{\mathbf{d}_i^H(f)\mathbf{\Phi}_{\mathbf{v}_i^{(n)}}^{-1}(f)\mathbf{d}_i(f)}$
  end
  $\mathbf{H}^{(n)}(f) = \begin{bmatrix} \mathbf{h}_1^{(n)H} & \mathbf{0}_{1\times 2} & \cdots & \mathbf{0}_{1\times 2} \\ \mathbf{0}_{1\times 2} & \mathbf{h}_2^{(n)H} & \cdots & \mathbf{0}_{1\times 2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0}_{1\times 2} & \mathbf{0}_{1\times 2} & \cdots & \mathbf{h}_{2^{N-n-1}}^{(n)H} \end{bmatrix}$
  $\mathbf{z}^{(n+1)}(f) = \mathbf{H}^{(n)}(f)\mathbf{z}^{(n)}(f)$
  $\mathbf{\Phi}_{\mathbf{v}^{(n+1)}}(f) = \mathbf{H}^{(n)}(f)\mathbf{\Phi}_{\mathbf{v}^{(n)}}(f)\mathbf{H}^{(n)H}(f)$
End

Outputs:
$Z_1^{(N)}(f) = \mathbf{z}^{(N)}(f)$, which is the estimate of the desired signal.
$\phi_{V_1^{(N)}}(f) = \mathbf{\Phi}_{\mathbf{v}^{(N)}}(f)$, which is the variance of the residual noise.

Because the conventional MVDR beamformer is the best linear filter that minimizes the variance of the residual noise, and both the conventional and multistage MVDR beamformers lie on the same hyperplane, $\mathbf{h}^H(f)\mathbf{d}(f)=1$, we deduce that the multistage MVDR beamformer is identical to the conventional one if their residual noise variances are equal.

In Table II, we show the complexity of the multistage MVDR beamformer, where "Mul," "Add," and "Div" denote multiplication, addition, and division, respectively. It is well known that the complexity of the conventional MVDR beamformer at every frequency is $\mathcal{O}(M^3)$ while that of the multistage MVDR beamformer is only $\mathcal{O}(M^2)$ as seen from Table II. So, the multistage MVDR beamformer is computationally much more efficient than its conventional counterpart, particularly when M is large.

V. PERFORMANCE MEASURES

Because microphone 1 is chosen as the reference, the performance measures for all approaches are defined with respect to this microphone.

We define the narrowband input signal-to-noise ratio (SNR) as the variance of the desired signal at frequency f over the variance of the noise at frequency f, i.e.,

$$\text{iSNR}(f)=\frac{\phi_{X_1}(f)}{\phi_{V_1}(f)}. \tag{36}$$

The broadband input SNR, which is a measure of the SNR across the entire frequency range, is defined as

$$\text{iSNR}=\frac{\int_f \phi_{X_1}(f)df}{\int_f \phi_{V_1}(f)df}. \tag{37}$$

The output SNR helps quantify the level of noise remaining at the beamformer output signal. With the conventional MVDR approach, the narrowband output SNR at frequency f is

$$\text{oSNR}(f)=\frac{\phi_{X_{\text{fd}}}(f)}{\phi_{V_{\text{rn}}}(f)}=\phi_{X_1}(f)\times\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f). \tag{38}$$

TABLE II. Computational complexity of the multistage MVDR beamformer.

Variable	Complexity	Number
$\mathbf{h}_i^{(n)}(f)$	8Mul + 3Add + 1Div	$\sum_{n=0}^{N-1}2^{N-n-1}=M-1$
$[\mathbf{\Phi}_{\mathbf{v}^{(n+1)}}(f)]_{i,j}$	6Mul + 3Add	$\frac{1}{2}\sum_{n=0}^{N-2}\left[(2^{N-n-1})^2+2^{N-n-1}\right]=\frac{1}{6}(M^2+3M-10)$
$\mathbf{h}_i^{(n)H}(f)\mathbf{z}_i^{(n)}(f)$	2Mul + 1Add	$\sum_{n=0}^{N-1}2^{N-n-1}=M-1$

It is easy to check that the broadband output SNR is

$$\text{oSNR}=\frac{\int_f \phi_{X_1}(f)df}{\int_f \phi_{V_{\text{rn}}}(f)df}. \tag{39}$$

The role of the beamformer is to produce a signal with a higher SNR in comparison with the observed signal. The amount of SNR improvement is measured by the so-called array gain Johnson and Dudgeon (1993). With the conventional MVDR filter, the narrowband array gain at frequency f is

$$\mathcal{A}(f)=\frac{\text{oSNR}(f)}{\text{iSNR}(f)}=\phi_{V_1}(f)\times\mathbf{d}^H(f)\mathbf{\Phi}_{\mathbf{v}}^{-1}(f)\mathbf{d}(f) \tag{40}$$

and the broadband array gain is

$$\mathcal{A}=\frac{\text{oSNR}}{\text{iSNR}}=\frac{\int_f \phi_{V_1}(f)df}{\int_f \phi_{V_{\text{rn}}}(f)df}. \tag{41}$$

From Benesty *et al.* (2008) and Souden *et al.* (2010) we know that

$$\text{oSNR}(f)\geq\text{iSNR}(f), \tag{42}$$

which leads to

$$\phi_{V_{\text{rn}}}(f)\leq\phi_{V_1}(f) \tag{43}$$

and

$$\mathcal{A}(f)\geq 1. \tag{44}$$

Integrating both sides of Eq. (43) over all frequencies, we get

$$\int_f \phi_{V_{\text{rn}}}(f)df\leq\int_f \phi_{V_1}(f)df, \tag{45}$$

which implies that

$$\mathcal{A}\geq 1 \tag{46}$$

and

$$\text{oSNR}\geq\text{iSNR}. \tag{47}$$

With the proposed multistage MVDR beamformer, the narrowband output SNR at frequency f and stage n is defined as

$$\begin{aligned}\text{oSNR}^{(n)}(f)&=\frac{\phi_{X_1}(f)}{\phi_{V_1^{(n)}}(f)}\\&=\phi_{X_1}(f)\times\mathbf{d}_1^H(f)\mathbf{\Phi}_{\mathbf{v}_1^{(n-1)}}^{-1}(f)\mathbf{d}_1(f).\end{aligned} \tag{48}$$

The broadband output SNR at stage n is

$$\text{oSNR}^{(n)} = \frac{\int_f \phi_{X_1}(f)df}{\int_f \phi_{V_1^{(n)}}(f)df}. \quad (49)$$

We easily deduce the narrowband and broadband array gains at stage n,

$$\mathcal{A}^{(n)}(f) = \frac{\text{oSNR}^{(n)}(f)}{\text{iSNR}(f)} = \phi_{V_1}(f) \times \mathbf{d}_1^H(f)\mathbf{\Phi}^{-1}_{\mathbf{v}_1^{(n-1)}}(f)\mathbf{d}_1(f), \quad (50)$$

$$\mathcal{A}^{(n)} = \frac{\text{oSNR}^{(n)}}{\text{iSNR}} = \frac{\int_f \phi_{V_1}(f)df}{\int_f \phi_{V_1^{(n)}}(f)df}. \quad (51)$$

It is clear that the multistage MVDR beamformer corresponds to N successive two-channel MVDR beamformers. The output of the $(n-1)$th beamformer is the input of the nth beamformer. Therefore the output SNR of the $(n-1)$th beamformer is the input SNR of the nth beamformer. From this observation and from Eq. (42), we then deduce that

$$\text{oSNR}^{(n)}(f) \geq \text{iSNR}^{(n)}(f) = \text{oSNR}^{(n-1)}(f). \quad (52)$$

Hence

$$\text{oSNR}^{(N)}(f) \geq \cdots \geq \text{oSNR}^{(1)}(f) \geq \text{oSNR}^{(0)}(f) = \text{iSNR}(f) \quad (53)$$

and

$$\mathcal{A}^{(N)}(f) \geq \cdots \geq \mathcal{A}^{(1)}(f) \geq 1. \quad (54)$$

We draw the same conclusions for the broadband measures, i.e.,

$$\text{oSNR}^{(N)} \geq \cdots \geq \text{oSNR}^{(1)} \geq \text{oSNR}^{(0)} = \text{iSNR}, \quad (55)$$

$$\mathcal{A}^{(N)} \geq \cdots \geq \mathcal{A}^{(1)} \geq 1. \quad (56)$$

We can express the narrowband output SNR at stage n as

$$\text{oSNR}^{(n)}(f) = \frac{\phi_{X_1}(f)}{\phi_{V_1^{(n-1)}}(f)} \times \frac{\phi_{V_1^{(n-1)}}(f)}{\phi_{V_1^{(n)}}(f)} = \text{oSNR}^{(n-1)}(f)\frac{\phi_{V_1^{(n-1)}}(f)}{\phi_{V_1^{(n)}}(f)}. \quad (57)$$

The previous equation leads to

$$\text{oSNR}^{(N)}(f) = \text{iSNR}(f)\frac{\phi_{V_1}(f)}{\phi_{V_1^{(N)}}(f)}. \quad (58)$$

As a result,

$$\mathcal{A}^{(N)}(f) = \frac{\phi_{V_1}(f)}{\phi_{V_1^{(N)}}(f)}. \quad (59)$$

It is also easy to verify that

$$\mathcal{A}^{(N)} = \frac{\int_f \phi_{V_1}(f)df}{\int_f \phi_{V_1^{(N)}}(f)df}. \quad (60)$$

The two previous expressions show how the the array gains depend on the variance of the noise at the reference microphone (before processing) and the variance of the residual noise at the last stage (after processing).

VI. PERFORMANCE STUDY

A. Performance in spatially uncorrelated noise

In this case, the correlation matrix of the noise at the nth stage is a diagonal one, i.e.,

$$\mathbf{\Phi}_{\mathbf{v}^{(n)}}(f) = \text{diag}[\phi_{V_1^{(n)}}(f), \phi_{V_2^{(n)}}(f), \ldots, \phi_{V_{2^{N-n}}^{(n)}}(f)], \quad (61)$$

where $\phi_{V_i^{(n)}}(f) \triangleq E[|V_i^{(n)}(f)|^2]$ is the variance of the residual noise of the ith channel at the nth stage. Its inverse, from Sec. IV, satisfies

$$\phi^{-1}_{V_i^{(n)}}(f) = \left|\frac{G_{2i-1}(f)}{G_i(f)}\right|^2 \phi^{-1}_{V_{2i-1}^{(n-1)}}(f) + \left|\frac{G_{2i}(f)}{G_i(f)}\right|^2 \phi^{-1}_{V_{2i}^{(n-1)}}(f). \quad (62)$$

From this equation, we deduce that

$$\begin{aligned}\phi^{-1}_{V_1^{(N)}}(f) &= \phi^{-1}_{V_1^{(N-1)}}(f) + \left|\frac{G_2(f)}{G_1(f)}\right|^2 \phi^{-1}_{V_2^{(N-1)}}(f) \\ &= \phi^{-1}_{V_1^{(N-2)}}(f) + \left|\frac{G_2(f)}{G_1(f)}\right|^2 \phi^{-1}_{V_2^{(N-2)}}(f) \\ &\quad + \left|\frac{G_3(f)}{G_1(f)}\right|^2 \phi^{-1}_{V_3^{(N-2)}}(f) + \left|\frac{G_4(f)}{G_1(f)}\right|^2 \phi^{-1}_{V_4^{(N-2)}}(f) \\ &\quad \vdots \\ &= \sum_{i=1}^{2^N} \left|\frac{G_i(f)}{G_1(f)}\right|^2 \phi^{-1}_{V_i^{(0)}}(f) \\ &= \sum_{i=1}^{2^N} \left|\frac{G_i(f)}{G_1(f)}\right|^2 \phi^{-1}_{V_i}(f). \end{aligned} \quad (63)$$

Clearly, the variances of the conventional and multistage MVDR beamformers are identical in the spatially uncorrelated noise and so are the two beamformers.

Substituting Eq. (63) into Eq. (59), we deduce the array gain

$$\begin{aligned}\mathcal{A}^{(N)}(f) &= \phi_{V_1}(f)\sum_{i=1}^{2^N} \left|\frac{G_i(f)}{G_1(f)}\right|^2 \phi^{-1}_{V_i}(f) \\ &= \frac{\phi_{V_1}(f)}{\phi_{X_1}(f)} \sum_{i=1}^{2^N} \frac{\phi_{X_i}(f)}{\phi_{V_i}(f)} \\ &= \frac{1}{\text{iSNR}_1(f)} \sum_{i=1}^{2^N} \text{iSNR}_i(f), \end{aligned} \quad (64)$$

where $\mathrm{iSNR}_i(f) = \phi_{X_i}(f)/\phi_{V_i}(f)$, $i = 1, 2, \ldots, M$ is the input SNR at the ith microphone. If the input SNRs at different sensors are the same, the array gain $\mathcal{A}^{(N)}(f)$ is M. Substituting Eq. (64) into Eq. (50), we find that

$$\mathrm{oSNR}^{(N)}(f) = \sum_{i=1}^{2^N} \mathrm{iSNR}_i(f), \tag{65}$$

which is the sum of the input SNRs over all sensors.

B. Performance in diffuse noise

If we are in the presence of the spherically isotropic (diffuse) noise field (Jacobsen, 1979; Goulding, 1989; Benesty *et al.*, 2007), its correlation matrix has the following form:

$$\mathbf{\Phi}_{\mathbf{v}}(f) = \phi_{V_1}(f)\mathbf{\Gamma}_{\mathrm{dn}}(f), \tag{66}$$

where $\mathbf{\Gamma}_{\mathrm{dn}}(f)$ is the pseudo-coherence matrix of the noise the (i, j)th element of which is

$$[\mathbf{\Gamma}_{\mathrm{dn}}(f)]_{i,j} = \mathrm{sinc}\left(2\pi f \delta_{i,j}/c\right) \triangleq \frac{\sin\left(2\pi f \delta_{i,j}/c\right)}{2\pi f \delta_{i,j}/c}, \tag{67}$$

with $\delta_{i,j}$ being the distance between the ith and jth sensors. The conditioning of this matrix depends on both the frequency and sensor spacing. To illustrate this, let us consider a uniform linear array consisting of eight microphones with a two-centimeter sensor spacing. Figure 2 plots the eigenvalues of $\mathbf{\Gamma}_{\mathrm{dn}}(f)$ as a function of frequency. Apparently, this matrix is very ill-conditioned in low frequencies because it has some very small eigenvalues in this case. So in low frequencies, the conventional MVDR beamformer may suffer from numerical problems, leading to significant sensor noise amplification. A very important metric to evaluate sensor noise amplification of a beamformer is the so-called white noise gain, which is defined as

$$\mathcal{A}_{\mathrm{wn}}(f) \triangleq \frac{1}{\mathbf{h}^H(f)\mathbf{h}(f)}, \tag{68}$$

where $\mathbf{h}(f)$ is the corresponding beamforming filter. If we assume that the desired source is in the farfield, in an anechoic environment, and arrives at the array from the end-fire direction, i.e., $\theta_{\mathrm{d}} = 0°$, we can write the steering vector according to Eq. (4). Then we can obtain both the array gain and white noise gain of the conventional and multistage MVDR beamformers by substituting Eq. (8) into Eqs. (40) and Eq. (68) and Eq. (34) into Eqs. (59) and (68). Figure 3 plots the array gain and white noise gain for both the conventional and multistage MVDR beamformers with a uniform linear array of eight sensors in diffuse noise. It is seen from Fig. 3 that in high frequencies the conventional MVDR beamformer achieves an array gain of approximately 18 dB. This corroborates with the theoretical analysis that the maximum gain of the MVDR beamformer in diffuse noise is M^2 (Uzkov, 1946). However, the array gain of the conventional MVDR beamformer is not stable in low frequencies mainly due to the ill-conditioning of the noise correlation matrix. While it achieves a great gain in reducing diffuse noise, the conventional MVDR beamformer also suffers from significant white noise amplification, which can be clearly seen in Fig. 3(b). In comparison, the multistage approach has less

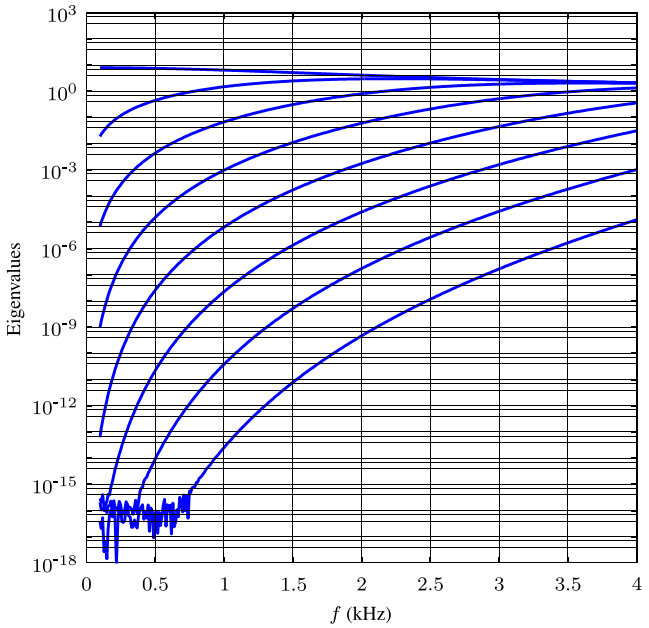

FIG. 2. (Color online) Eigenvalues of the diffuse noise pseudo-coherence matrix as a function of frequency with an eight-element uniform linear array ($\delta = 2$ cm).

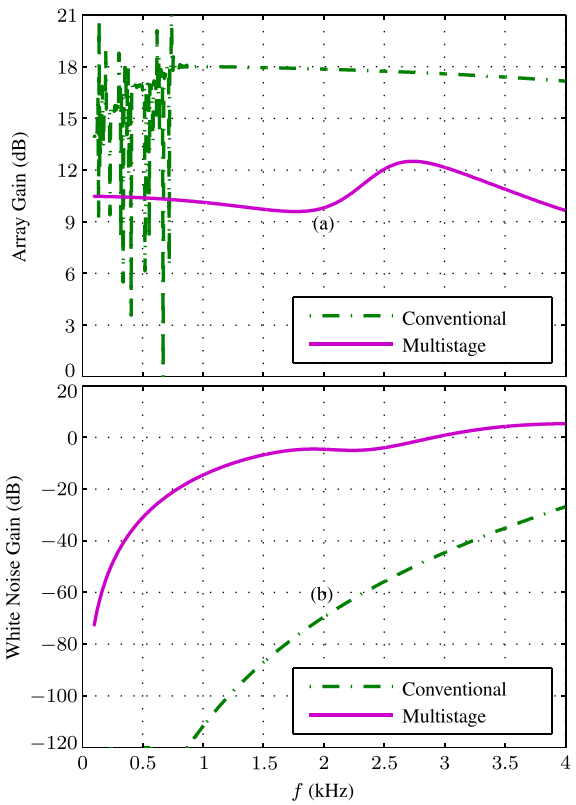

FIG. 3. (Color online) Performance of the conventional and multistage MVDR beamformers with an eight-element uniform linear array ($\delta = 2$ cm): (a) Array gain and (b) white noise gain.

array gain, but it is significantly better in dealing with white noise amplification.

The performance of the MVDR beamformer, no matter how it is implemented, depends on the source incidence angle as shown in Pan *et al.* (2014). Figure 4 plots the array gain as a function of the frequency and source incidence angle of both the conventional and multistage MVDR beamformers with an eight-sensor uniform linear array in diffuse noise. It is clearly seen that the maximum gain is achieved when the source is in the endfire directions while minimum gain occurs when the source is incident from the broadside, i.e., $\theta_d = 90°$. When the source is at the broadside, the array gain of the conventional MVDR beamformer does not change much with frequency. However, the multistage MVDR beamformer has less array gain as the frequency decreases. Indeed, when $\theta_d = 90°$, the array gain of the multistage algorithm can be written as

$$\mathcal{A}^{(N)}(f) = \frac{M^2}{\sum_{m=-(M-1)}^{M-1} (M - |m|)\mathrm{sinc}(2\pi f m\delta/c)}. \tag{69}$$

One can check that

$$\mathcal{A}^{(N)}(f) = \begin{cases} 1, & f \to 0 \\ M, & f \to \infty. \end{cases} \tag{70}$$

So there is not much SNR improvement with respect to diffuse noise when the frequency is very low. The underlying reason is that the multistage MVDR beamformer attempts to improve the white noise amplification problem by sacrificing some performance gain in dealing with diffuse noise, as shown in Fig. 5, where the white noise gain of the multistage MVDR beamformer is much larger than that of its conventional counterpart.

Like the array gain, the white noise gains of the conventional and multistage MVDR beamformers are also a function of frequency and source incidence angle as seen in Fig. 5. Apparently, the white noise gain of the conventional MVDR beamformer decreases dramatically with frequency for a given source incidence angle; but at a few angles, the white noise gain is much larger than that at most other frequencies. This phenomenon is explained in the Appendix. In comparison, the white noise gain of the multistage MVDR beamformer is quite flat and does not change much with the source incidence angle.

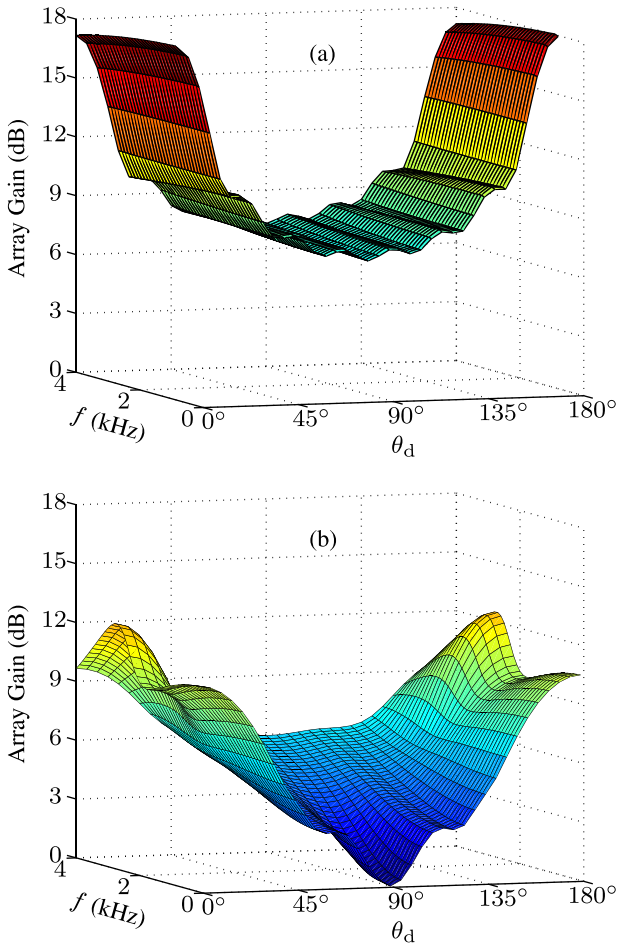

FIG. 4. (Color online) Array gain of the conventional and multistage MVDR beamformers as a function of frequency and source incidence angle with an eight-element uniform linear array ($\delta = 2$ cm): (a) Conventional beamformer and (b) multistage beamformer.

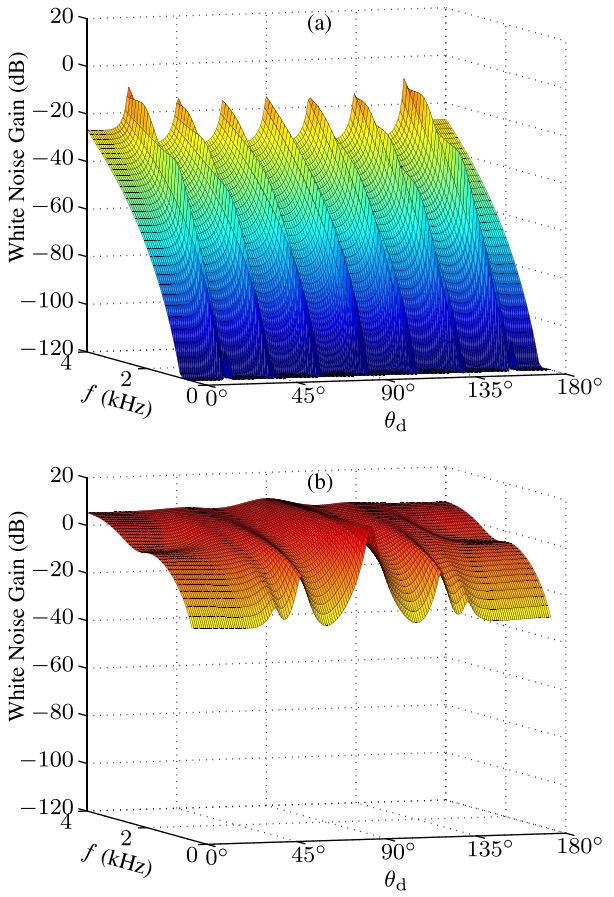

FIG. 5. (Color online) White noise gain of the conventional and multistage MVDR beamformers as a function of frequency and source incidence angle with an with-element uniform linear array ($\delta = 2$ cm): (a) Conventional beamformer and (b) multistage beamformer.

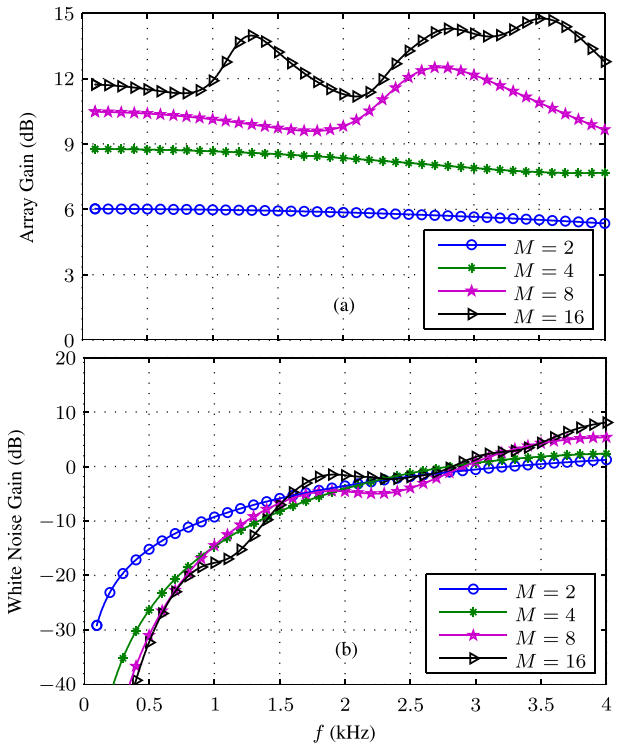

FIG. 6. (Color online) Performance of the multistage MVDR beamformer with a uniform linear array ($\delta = 2$ cm) for different values of M.

C. Performance with different number of sensors

The number of sensors plays an important role in the array performance. Figure 6 plots the array gain and white noise gain of the multistage MVDR beamformer with a uniform linear array, as a function of frequency, for different values of M in diffuse noise and $\theta_d = 0°$. It is seen that the array gain increases with the number of sensors while the white noise gain seems not to change much. This indeed shows the robustness of the multistage MVDR beamformer from another perspective.

It is seen that when the number of sensors is increased from two to four, the array gain is improved by approximately 3 dB. However, when the number of sensors is large, doubling the number of sensors gives a gain improvement less than 3 dB. To explain this, we plot in Fig. 7 the level distribution of the residual noise at different stages. It can be seen that the energy of the noise concentrates more and more on the desired source direction as the stage is increased, and as a result, it is more difficult to reduce diffuse noise from one stage to the next.

D. Performance with different subarray structures

In both the theoretical analysis and previous performance study, we divided the microphone array into subarrays of two microphones each. This idea can be easily generalized to the general case where each subarray has L microphones with $L \leq M$. When $L = M$, the multistage approach degenerates to the conventional MVDR beamformer. Generally, as the value of L increases, the array gain of the multistage MVDR beamformer gets closer to that of the conventional MVDR beamformer; but it has more white noise amplification just like the conventional MVDR beamformer.

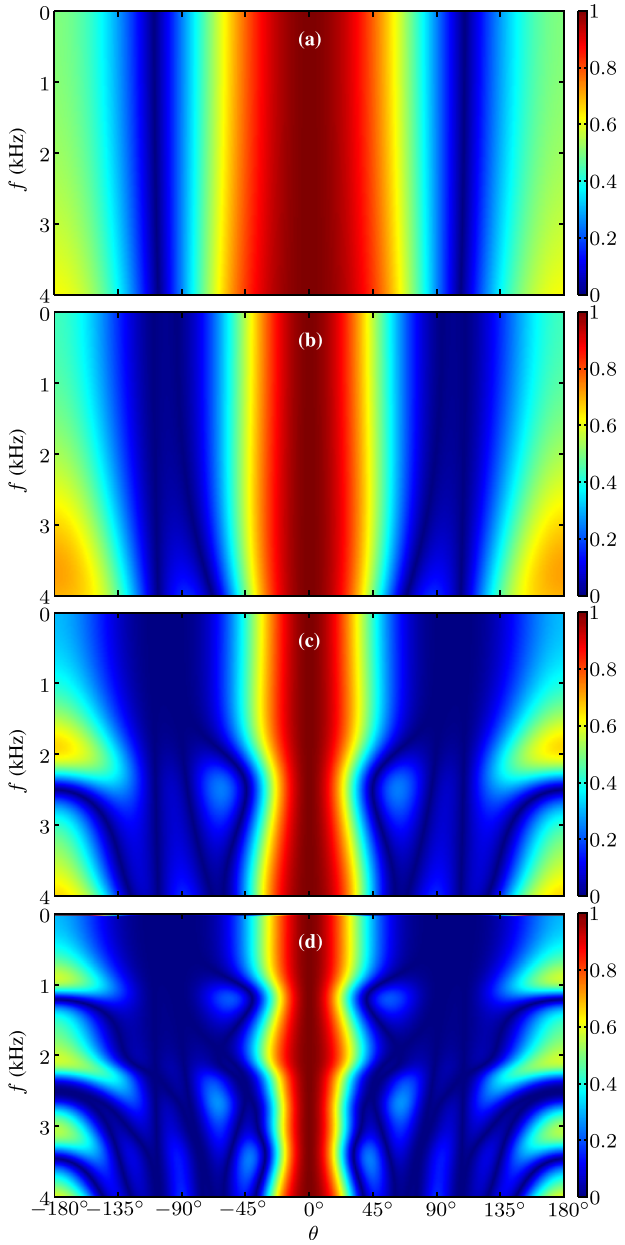

FIG. 7. (Color online) The noise spatial distribution of the multistage MVDR beamformer with a uniform linear array, 16 sensors, $\delta = 2$ cm, and $\theta_d = 0°$: (a) First stage, (b) second stage, (c) third stage, and (d) fourth stage.

To illustrate this, we use a uniform linear array with 16 microphones and spacing of 2 cm. We consider two cases: (1) Each subarray consists of two microphones and (2) each subarray consists of four microphones. The results are plotted in Fig. 8. It is clearly seen that the array gain with respect to diffuse noise increases if the subarray uses more microphones; but its has more white noise amplification as compared to the case with less microphones.

VII. CONCLUSIONS

In this paper, we developed a multistage MVDR beamformer. Unlike the conventional MVDR beamformer that forms the beamforming filter using all the sensors at once, this multistage approach first divides the microphone array of M sensors into $M/2$ subarrays with each subarray having only

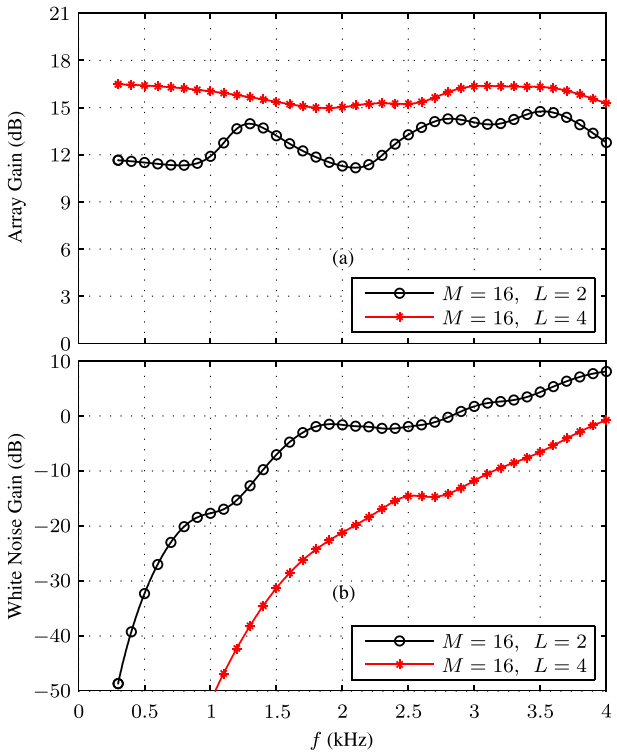

FIG. 8. (Color online) Performance of the multistage MVDR beamformer with a uniform linear array ($\delta = 2$ cm) and two subarray structures.

two microphones. A two-channel MVDR beamformer is performed with every subarray. The $M/2$ subarrays' outputs are then treated as the inputs of $M/4$ subarrays of two channels in the next stage. Similarly, a two-channel MVDR beamformer is performed with each subarray in the second stage. This process is repeated till the last stage that has only a single output. Through both theoretical analysis and simulations, we showed that this multistage MVDR beamformer has following appealing properties. First, the array performance gradually increases from one stage to the next. Second, after the final stage, the performance of this multistage approach is identical to that of the conventional MVDR beamformer in spatially white noise. Third, it is much more robust than the conventional MVDR beamformer in diffuse noise (it has significant higher white noise gain). Moreover, its complexity is an order of magnitude smaller than that of the conventional MVDR beamformer. We also showed that the basic principle in this paper can be easily generalized to the case where every subarray has more than two microphones. In this case, the performance behavior of the multistage MVDR beamformer gets closer to the conventional one as the number of sensors in each subarray increases; but the robustness decreases while the complexity increases.

ACKNOWLEDGMENT

This work was supported in part by the NSFC "Distinguished Young Scientists Fund" under Grant No. 61425005.

APPENDIX

In this appendix, we show that for a uniform linear array with M sensors in diffuse noise, the white noise gain at any particular frequency is a function of the source incidence angle θ_d, and it generally has $M-1$ local maxima for $\theta_d \in [0, 180°)$.

The white noise gain of the conventional MVDR filter in diffuse noise can be written into the following form based on Eqs. (8), (66), and (68):

$$\mathcal{A}_{wn}(f) = \frac{|\mathbf{d}^H(f)\boldsymbol{\Gamma}_{dn}^{-1}(f)\mathbf{d}(f)|^2}{\mathbf{d}^H(f)\boldsymbol{\Gamma}_{dn}^{-2}(f)\mathbf{d}(f)}. \tag{A1}$$

Using the eigenvalue decomposition, $\boldsymbol{\Gamma}_{dn}(f)$ can be decomposed as

$$\boldsymbol{\Gamma}_{dn}(f) = \mathbf{U}\boldsymbol{\Lambda}\mathbf{U}^H, \tag{A2}$$

where

$$\boldsymbol{\Lambda} = \text{diag}\,[\lambda_1, \lambda_2, \ \ldots, \ \lambda_M] \tag{A3}$$

is a diagonal matrix consisting of all the eigenvalues of $\boldsymbol{\Gamma}_{dn}(f)$ with $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_M > 0$, and

$$\mathbf{U} = \begin{bmatrix}\mathbf{u}_1 & \mathbf{u}_2 & \cdots & \mathbf{u}_M\end{bmatrix} \tag{A4}$$

consists of the corresponding eigenvectors. Using these eigenvectors, one can rewrite the steering vector $\mathbf{d}(f)$ into the following form:

$$\mathbf{d}(f) = \sum_{m=1}^{M} \sqrt{\lambda_m}\zeta_m(\cos\theta_d)e^{\jmath\phi_m}\mathbf{u}_m, \tag{A5}$$

where

$$\zeta_m(\cos\theta_d) = \frac{1}{\sqrt{\lambda_m}}\mathbf{u}_m^H\mathbf{d}(f)e^{-\jmath\phi_m}, \tag{A6}$$

is called the eigen pattern of the beamformer, and ϕ_m is a phase to make $\zeta_m(\cos\theta_d)$ a real function. It follows then that

$$\begin{aligned}\mathcal{A}_{wn}(f) &= \frac{|[\mathbf{U}^H\mathbf{d}(f)]^H\boldsymbol{\Lambda}^{-1}[\mathbf{U}^H\mathbf{d}(f)]|^2}{[\mathbf{U}^H\mathbf{d}(f)]^H\boldsymbol{\Lambda}^{-2}[\mathbf{U}^H\mathbf{d}(f)]} \\ &= \frac{\left|\sum_{m=1}^{M}|\zeta_m(\cos\theta_d)|^2\right|^2}{\sum_{m=1}^{M}\lambda_m^{-1}|\zeta_m(\cos\theta_d)|^2} \\ &= \lambda_M\frac{\left|\sum_{m=1}^{M}|\zeta_m(\cos\theta_d)|^2\right|^2}{|\zeta_m(\cos\theta_d)|^2 + \sum_{m=1}^{M-1}\frac{\lambda_M}{\lambda_m}|\zeta_m(\cos\theta_d)|^2}.\end{aligned} \tag{A7}$$

Because λ_M is the smallest eigenvalue of $\boldsymbol{\Gamma}_{dn}$ and it is much smaller than the other eigenvalues when M is reasonably large, we generally have

$$|\zeta_M(\cos\theta_d)|^2 \gg \sum_{m=1}^{M-1}\frac{\lambda_M}{\lambda_m}|\zeta_m(\cos\theta_d)|^2, \tag{A8}$$

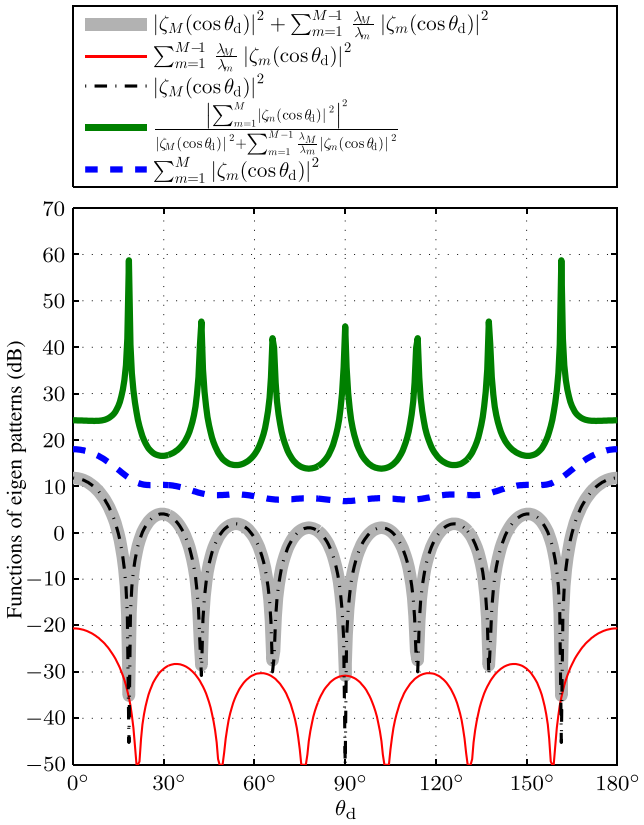

FIG. 9. (Color online) Illustration of the white noise gain of the conventional MVDR beamformer with a uniform linear array in diffuse noise: $M = 8$, $\delta = 2\,\text{cm}$, $f = 1\,\text{kHz}$, and $10\log_{10}(\lambda_M) \approx -136\,\text{dB}$.

which can be seen from Fig. 9. Consequently, with a given frequency, the white noise gain of the conventional MVDR beamformer in diffuse noise reaches its local maxima when $\zeta_M(\cos\theta_d)$ reaches its local minima, as illustrated in Fig. 9. This also corroborates the results shown in Fig. 5.

Benesty, J., Chen, J., and Huang, Y. (**2008**). *Microphone Array Signal Processing* (Springer-Verlag, Berlin, Germany), pp. 1–232.

Benesty, J., Sondhi, M. M., and Huang, Y., eds. (**2007**). *Springer Handbook of Speech Processing* (Springer-Verlag, Berlin, Germany), pp. 1–1176.

Brandstein, M., and Ward, D. B., eds. (**2001**). *Microphone Arrays: Signal Processing Techniques and Applications* (Springer-Verlag, Berlin, Germany), pp. 1–398.

Breed, B. R., and Strauss, J. (**2004**). "A short proof of the equivalence of LCMV and GSC beamforming," IEEE Signal Process. Lett. **9**, 168–169.

Capon, J. (**1969**). "High resolution frequency-wavenumber spectrum analysis," Proc. IEEE **57**, 1408–1418.

Carlson, B. D. (**1988**). "Covariance matrix estimation errors and diagonal loading in adaptive arrays," IEEE Trans. Aerosp. Electron. Syst. **24**, 397–401.

Cox, H., and Pitre, R. (**1998**). "Robust DMR and multi-rate adaptive beamforming," in *Proc. Thirty-First Asilomar Conference on Signals, Systems and Computers*, pp. 920–924.

Frost, O. L. III (**1972**). "An algorithm for linearly constrained adaptive array processing," Proc. IEEE **60**, 926–935.

Gannot, S., Burshtein, D., and Weinstein, E. (**2001**). "Signal enhancement using beamforming and nonstationarity with applications to speech," IEEE Trans. Signal Process. **49**, 1614–1626.

Goulding, M. (**1989**). "Speech enhancement for mobile telephony [microform]," M.A.Sc. thesis, Simon Fraser University.

Griffiths, L. J., and Jim, C. W. (**1982**). "An alternative approach to linearly constrained adaptive beamforming," IEEE Trans. Antennas Propagat. **30**, 27–34.

Jacobsen, F. (**1979**). "The diffuse sound field: Statistical considerations concerning the reverberant field in the steady state," Report (Acoustics Laboratory, Technical University of Denmark).

Johnson, D. H., and Dudgeon, D. E. (**1993**). *Array Signal Processing—Concepts and Techniques* (Prentice Hall, Englewood Cliffs, NJ), pp. 1–512.

Kogon, S. M. (**2004**). "Eigenvectors, diagonal loading and white noise gain constraints for robust adaptive beamforming," in *Proc. the Thirty-Seventh Asilomar Conference on Signals, Systems and Computers*, Vol. 2, pp. 1853–1857.

Lacoss, R. T. (**1971**). "Data adaptive spectral analysis methods," Geophysics **36**, 661–675.

Li, J., Stoica, P., and Wang, Z. (**2003**). "On robust Capon beamforming and diagonal loading," IEEE Trans. Signal Process. **51**, 1702–1715.

Pados, D. A., and Karystinos, G. N. (**2001**). "An iterative algorithm for the computation of the MVDR filter," IEEE Trans. Signal Process. **49**, 290–300.

Pan, C., Chen, J., and Benesty, J. (**2014**). "Performance study of the MVDR beamformer as a function of the source incidence angle," IEEE Trans. Audio Speech Lang. Process. **22**, 67–79.

Souden, M., Benesty, J., and Affes, S. (**2010**). "On the global output SNR of the parameterized frequencydomain multichannel noise reduction Wiener filter," IEEE Signal Process. Lett. **17**, 425–428.

Uzkov, A. I. (**1946**). "An approach to the problem of optimum directive antenna design," Comptes Rendus (Doklady) de l'Academie des Sciences de l'URSS **LIII**, 35–38.