

Performance Study of the MVDR Beamformer as a Function of the Source Incidence Angle

Chao Pan, *Student Member, IEEE*, Jingdong Chen, *Senior Member, IEEE*, and Jacob Benesty

Abstract—Linear microphone arrays combined with the minimum variance distortionless response (MVDR) beamformer have been widely studied in various applications to acquire desired signals and reduce the unwanted noise. Most of the existing array systems assume that the desired sources are in the broadside direction. In this paper, we study and analyze the performance of the MVDR beamformer as a function of the source incidence angle. Using the signal-to-noise ratio (SNR) and beampattern as the criteria, we investigate its performance in four different scenarios: spatially white noise, diffuse noise, diffuse-plus-white noise, and point-source-plus-white noise. The results demonstrate that the optimal performance of the MVDR beamformer occurs when the source is in the endfire directions for diffuse noise and point-source noise while its SNR gain does not depend on the signal incidence angle in spatially white noise. This indicates that most current systems may not fully exploit the potential of the MVDR beamformer. This analysis does not only help us better understand this algorithm, but also helps us design better array systems for practical applications.

Index Terms—Beamforming, diffuse noise, microphone arrays, minimum variance distortionless response (MVDR) beamformer, noise reduction, spatially white noise, speech enhancement.

I. INTRODUCTION

MICROPHONE arrays, which consist of sets of acoustic sensors that are spatially arranged in specific geometries, have a wide range of applications. In such a system, each sensor collects signals from sources in its own field of view and, therefore, the array outputs contain the signal of interest, noise, interference, and also the propagation information that is represented by the acoustic impulse responses from the radiating sources to the microphones. By applying filters to the outputs of the sensors and combining the results together, various functionalities can be implemented including but not limited to: localizing and tracking the sound sources, extracting the signal of interest, suppressing ambient noise, and separating different sound sources. This spatial filtering process, called beamforming, plays a critical role in a microphone array system and

controls how well the system works in practical environments. As a result, tremendous efforts have been devoted to the design of beamformers and many have been developed over the last four decades. Broadly, those algorithms can be divided into two categories: fixed beamformers and adaptive ones.

Fixed beamformers have static (once designed) filter coefficients and, ideally, signal-independent spatial responses. The simplest one is the delay-and-sum beamformer that was originally developed in the sonar and radar applications [1]–[4]. The basic idea is to delay each sensor output by a proper amount of time so that the signal components from the desired source are synchronized across all sensors. These delayed signals are then weighted and summed together. Since they add up together coherently, the desired signal components are reinforced. In contrast, the other sources and noise are suppressed or even eliminated as they are added together destructively. This beamformer, although it serves as the basis for many advanced algorithms, is only good for processing narrowband signals. When applied to broadband signals like speech, it generally produces different spatial responses at different frequencies, leading to either distortion of the desired signal or artifacts in the residual noise and interference [5]. There are two ways to circumvent this issue. The first one is to use harmonically nested subarrays where every subarray is designed for operating at a single frequency [6], [7]. Through adjusting the number of sensors and the spacing between them, all the subarrays can be designed to have a similar spatial response. However, harmonically-nested subarrays are large in size and require a great number of microphones, which prevent them from practical usage. The second way is based on the use of subband techniques, i.e., still a single array is used, but the signals are decomposed into subbands. In each subband, some constraints are applied in the beamformer design so that the spatial responses would be similar across all the subbands [8]. Theoretically, this subband approach is equivalent to the widely known filter-and-sum framework proposed by Frost in the early 1970s [9], which first applies a finite-impulse-response (FIR) filter to each microphone output and then sums all the filtered signals together. In addition to being able to produce a broadband spatial response, the filter-and-sum beamformer is more effective in suppressing noise and interference as compared to the simple delay-and-sum technique since it can produce more nulls. Consequently, this beamformer is practically more useful and has been intensively studied in the literature [4], [10]–[22]. In the design of a fixed filter-and-sum beamformer, the filter coefficients can be computed using a least-squares (LS) filter design method [21].

Fixed beamformers are data-independent spatial filters. In their design, the noise field is not known and the isotropic

Manuscript received May 03, 2013; revised September 08, 2013; accepted September 15, 2013. Date of publication September 24, 2013; date of current version November 13, 2013. The associate editor coordinating the review of this manuscript and approving it for publication was Prof. Boaz Rafaely.

C. Pan is with the Research Center of Intelligent Acoustics and Immersive Communications, Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China (e-mail: panchao2nwpw@mail.nwpu.edu.cn).

J. Chen is with Northwestern Polytechnical University, Xi'an, Shaanxi 710072, China (e-mail: jingdongchen@ieee.org).

J. Benesty is with INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada (e-mail: benesty@emt.inrs.ca).

Color versions of one or more of the figures in this paper are available online at <http://ieeexplore.ieee.org>.

Digital Object Identifier 10.1109/TASL.2013.2283104

model, which is a first-order approximation of most of the real noise fields, is commonly used. If this assumption matches the acoustic condition in which the array operates, the array may produce reasonably good performance. However, such an assumption in general does not hold in practice and therefore a fixed beamformer is at best suboptimal in real applications. For example, if noise comes from a point sound source, ideally the beamformer should be able to put a null in that direction and therefore completely remove the noise. However, this can only be done by changing the beamforming filter coefficients, leading to the class of adaptive beamformers. The fundamental principle of adaptive beamforming is to track the statistics of the surrounding noise field and adaptively search for the optimum location of the nulls that can most significantly reduce noise under the constraint that the desired speech signal is not distorted at the beamformer's output. The most representative adaptive beamformer is the linearly constrained minimum variance (LCMV) algorithm. It estimates the beamforming coefficients in an adaptive way by minimizing the variance of the residual noise and interference while enforcing a set of linear constraints to ensure that the desired signals are not distorted. This basic idea was first proposed by Frost [9] and, therefore, the LCMV beamformer is sometimes called the Frost beamformer. Since it involves matrix inversions, the direct implementation of an LCMV beamformer is in general computationally expensive and may suffer numerical problems as well. One variant of the LCMV beamformer is the so-called generalized sidelobe canceller (GSC) proposed by Griffiths and Jim [11]. The GSC transforms the LCMV algorithm from a constrained optimization problem into an unconstrained form. Theoretically, the GSC and LCMV beamformers are the same while the GSC can lower the computational cost by forcing the constraint into the front-end of the array processing [10], [17]. The LCMV beamformer is theoretically very appealing since it is an optimal spatial filter. But in practice, when an LCMV beamformer is used in a reverberant environment, some knowledge of the source propagation such as the impulse responses need to be known, which limits the application of this algorithm. One way to relax the *a priori* information is to assume that only the direction of the speech source needs to be known. In this case, the LCMV structure degenerates to the minimum variance distortionless response (MVDR) beamformer, which was originally developed by Capon (accordingly, it is often called the Capon beamformer) [23]. Although it can be viewed as a particular case of the LCMV method, the MVDR beamformer is in general more practical in terms of implementation as it requires less *a priori* knowledge and, therefore, it has attracted a significant amount of attention in the field of acoustic signal processing. Indeed, tremendous efforts have been devoted to the design, implementation, and analysis of this beamformer. Nowadays, many experimental platforms and even some commercial products start to use this beamformer in real-world applications to record speech signals for voice communication.

However, a very important question regarding the MVDR beamformer remains unclear, i.e., given an array system and an application scenario, how should one configure the system so that the MVDR beamformer can achieve its best performance in

terms of signal enhancement and noise and interference reduction? This paper is dedicated to answering this question. Based on the use of a linear microphone array, we investigate how the MVDR beamformer works in four different scenarios: spatially white noise, diffuse noise, diffuse-plus-white noise, and point-source-plus-white noise. The results demonstrate that the MVDR beamformer may differ significantly in its performance between endfire and broadside directions. When there is either diffuse or point-source noise in the acoustic field, which happens in most real acoustic environments, the optimal SNR gain occurs when the desired source is in the endfire directions. But the SNR gain does not depend on the signal incidence angle if only spatially white noise is present. This analysis does not only help us better understand the MVDR beamformer, but also helps us design better array systems in practical applications. Most importantly, it tells that a linear array should be configured in such a way that the desired source is in the endfire directions in order that the adaptive MVDR beamformer achieves its best performance.

The remainder of this paper is organized as follows. In Section II, we briefly describe the signal model used in this paper. We then discuss the concept of beamforming and the performance measures in Section III. Section IV derives the MVDR beamformer. In Section V, we analyze the SNR gain and beampattern of the MVDR beamformer in different noise environments. We then present some evaluations of the MVDR beamformer in simulated acoustic environments in Section VI. Finally, some conclusions are drawn in Section VII.

II. SIGNAL MODEL

In real acoustic environments, signals picked up by microphones consist of the desired signal, noise, and signal components due to the multipath effect. However, to make the exposition easy and clear, we first neglect the multipath effect and consider a simple signal model where a desired speech source (plane wave) propagates in an anechoic acoustic environment and impinges on a uniform linear array consisting of M omnidirectional microphones, as shown in Fig. 1. Let us choose the first microphone as the reference point, the signal received by the m th microphone ($m = 1, 2, \dots, M$) can then be written as [21]

$$y_m(t) = x_m(t) + v_m(t) = x(t - \tau_m) + v_m(t), \quad (1)$$

where $y_m(t)$, $x_m(t)$, and $v_m(t)$ are the noisy, clean speech, and noise signals, respectively, captured by the m th microphone at time t , $\tau_m = (m-1)\tau_0 \cos \theta_d$ is the relative time delay between the m th microphone and the reference sensor, $\tau_0 = \delta/c$ with δ being the spacing between two neighboring sensors and c being the speed of sound in air, i.e., $c = 340$ m/s, θ_d is the incidence angle of the desired sound source, and $x(t) = x_1(t)$ is the clean signal received at the reference microphone. All signals are considered to be zero-mean, real, and broadband. Furthermore, the noise signals $v_m(t)$, $m = 1, 2, \dots, M$, are assumed to be uncorrelated with the clean signals $x_m(t)$, $m = 1, 2, \dots, M$.

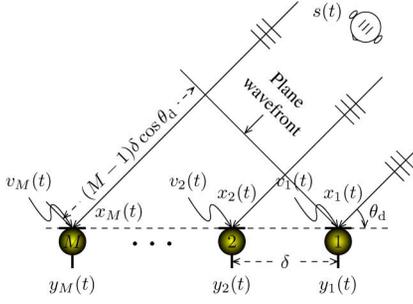


Fig. 1. Illustration of a uniform linear microphone array system, where M is the number of microphones, δ is the microphone spacing, and θ_d is the incidence angle of the desired source which is located in the far field.

To make the processing efficient, we work in the frequency domain. In this domain, the signal model given in (1) is written as

$$\begin{aligned} Y_m(\omega) &= X_m(\omega) + V_m(\omega) \\ &= e^{-j\omega\tau_m} X(\omega) + V_m(\omega) \\ &= e^{-j(m-1)\omega\tau_0 \cos \theta_d} X(\omega) + V_m(\omega), \end{aligned} \quad (2)$$

where $Y_m(\omega)$, $X_m(\omega)$, $V_m(\omega)$, and $X(\omega)$ are the Fourier transforms of $y_m(t)$, $x_m(t)$, $v_m(t)$, and $x(t)$, respectively, j is the imaginary unit, i.e., $j^2 = -1$, $\omega = 2\pi f$ is the angular frequency, and $f(> 0)$ denotes the temporal frequency. We can rearrange (2) into the following vector form:

$$\begin{aligned} \mathbf{y}(\omega) &\triangleq [Y_1(\omega) Y_2(\omega) \cdots Y_M(\omega)]^T \\ &= \mathbf{d}_{\theta_d}(\omega) X(\omega) + \mathbf{v}(\omega), \end{aligned} \quad (3)$$

where the superscript T is the transpose operator,

$$\mathbf{d}_{\theta_d}(\omega) \triangleq \left[1 e^{-j\omega\tau_0 \cos \theta_d} \cdots e^{-j(M-1)\omega\tau_0 \cos \theta_d} \right]^T, \quad (4)$$

and the noise signal vector, $\mathbf{v}(\omega)$, is defined in a similar manner to $\mathbf{y}(\omega)$.

By assumption, the signals $X(\omega)$ and $V_m(\omega)$ are zero-mean and uncorrelated with each other. Therefore, the correlation matrix of $\mathbf{y}(\omega)$ is

$$\begin{aligned} \Phi_{\mathbf{y}}(\omega) &\triangleq E[\mathbf{y}(\omega)\mathbf{y}^H(\omega)] \\ &= \phi_X(\omega)\mathbf{d}_{\theta_d}(\omega)\mathbf{d}_{\theta_d}^H(\omega) + \Phi_{\mathbf{v}}(\omega), \end{aligned} \quad (5)$$

where $E[\cdot]$ denotes mathematical expectation, the superscript H denotes the conjugate-transpose operator, $\phi_X(\omega) \triangleq E[|X(\omega)|^2]$ is the variance of $X(\omega)$, and $\Phi_{\mathbf{v}}(\omega) \triangleq E[\mathbf{v}(\omega)\mathbf{v}^H(\omega)]$ is the correlation matrix of $\mathbf{v}(\omega)$, which is assumed to be of full rank, i.e., equal to M .

Let us also introduce the pseudo-coherence matrix of $\mathbf{v}(\omega)$, which is defined as

$$\Gamma_{\mathbf{v}}(\omega) \triangleq \frac{\Phi_{\mathbf{v}}(\omega)}{\phi_{V_1}(\omega)}, \quad (6)$$

where $\phi_{V_1}(\omega) \triangleq E[|V_1(\omega)|^2]$ is the variance of $V_1(\omega)$. This Hermitian matrix is used in the next sections to study beamforming and analyze the performance measures.

III. BEAMFORMING AND PERFORMANCE MEASURES

With the frequency-domain signal model given in (3), the objective of beamforming is to recover the clean speech signal, $X(\omega)$, given the observation signals, $Y_m(\omega)$, $m = 1, 2, \dots, M$. This can be achieved by applying a complex weight to $Y_m(\omega)$ and then summing all the M weighted signals together, i.e.,

$$\begin{aligned} Z(\omega) &= \sum_{m=1}^M H_m^*(\omega) Y_m(\omega) \\ &= \mathbf{h}^H(\omega) \mathbf{y}(\omega) \\ &= \mathbf{h}^H(\omega) \mathbf{d}_{\theta_d}(\omega) X(\omega) + \mathbf{h}^H(\omega) \mathbf{v}(\omega), \end{aligned} \quad (7)$$

where the superscript $*$ is the complex conjugate operator, $Z(\omega)$ is an estimate of $X(\omega)$, and

$$\mathbf{h}(\omega) = [H_1(\omega) H_2(\omega) \cdots H_M(\omega)]^T \quad (8)$$

is the beamforming filter. Then, the objective of beamforming is to design an optimal filter, $\mathbf{h}(\omega)$, such that $Z(\omega)$ is a ‘‘good’’ estimate of $X(\omega)$. The performance is generally evaluated with two important metrics: beampattern and SNR gain.

A. Beampattern

Each beamformer has a pattern of directional sensitivity, i.e., it has different sensitivities for sounds arriving from different directions. The beampattern, also called the directivity pattern, describes the sensitivity of the beamformer to a plane wave (source signal) impinging on the array from the direction θ . For a uniform linear array with M sensors, the steering vector [2] is

$$\mathbf{d}(\omega) \triangleq \left[1 e^{-j\omega\tau_0 \cos \theta} \cdots e^{-j(M-1)\omega\tau_0 \cos \theta} \right]^T. \quad (9)$$

Therefore, the beampattern is defined as

$$\begin{aligned} \mathcal{B}[\mathbf{h}(\omega), \theta] &\triangleq \mathbf{d}_{\theta}^H(\omega) \mathbf{h}(\omega) \\ &= \sum_{m=1}^M H_m(\omega) e^{j(m-1)\omega\tau_0 \cos \theta}. \end{aligned} \quad (10)$$

Note that the beampattern of a linear array is rotationally symmetric with respect to a line through all the microphone positions, which can be easily checked from (9) and (10).

B. SNR Gain

With the signal model in (2), the input SNR of an array is defined as the SNR at the reference sensor, i.e.,

$$\text{iSNR}(\omega) \triangleq \frac{\phi_X(\omega)}{\phi_{V_1}(\omega)}. \quad (11)$$

With the beamformer's output given in (7), the output SNR is given by

$$\begin{aligned} \text{oSNR}[\mathbf{h}(\omega)] &= \frac{E \left[|\mathbf{h}^H(\omega) \mathbf{d}_{\theta_d}(\omega) X(\omega)|^2 \right]}{E \left[|\mathbf{h}^H(\omega) \mathbf{v}(\omega)|^2 \right]} \\ &= \text{iSNR}(\omega) \cdot \frac{|\mathbf{h}^H(\omega) \mathbf{d}_{\theta_d}(\omega)|^2}{\mathbf{h}^H(\omega) \mathbf{\Gamma}_v(\omega) \mathbf{h}(\omega)}, \end{aligned} \quad (12)$$

which depends on the input SNR, the signal incidence angle, the beamforming filter, as well as the pseudo-coherence matrix of the noise signal.

The definition of the SNR gain is easily derived from (12) and (11), i.e.,

$$\mathcal{G}[\mathbf{h}(\omega)] = \frac{\text{oSNR}[\mathbf{h}(\omega)]}{\text{iSNR}(\omega)} = \frac{|\mathbf{h}^H(\omega) \mathbf{d}_{\theta_d}(\omega)|^2}{\mathbf{h}^H(\omega) \mathbf{\Gamma}_v(\omega) \mathbf{h}(\omega)}. \quad (13)$$

For any two vectors $\mathbf{h}(\omega)$ and $\mathbf{d}_{\theta_d}(\omega)$, we have the following inequality:

$$|\mathbf{h}^H(\omega) \mathbf{d}_{\theta_d}(\omega)|^2 \leq [\mathbf{h}^H(\omega) \mathbf{\Gamma}_v(\omega) \mathbf{h}(\omega)] \cdot [\mathbf{d}_{\theta_d}^H(\omega) \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega)], \quad (14)$$

with equality if and only if $\mathbf{h}(\omega) \propto \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega)$. Applying this inequality to (13), we deduce an upper bound of the SNR gain:

$$\begin{aligned} \mathcal{G}[\mathbf{h}(\omega)] &\leq \mathbf{d}_{\theta_d}^H(\omega) \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega) \\ &\leq \text{tr} \left[\mathbf{\Gamma}_v^{-1}(\omega) \right] \text{tr} \left[\mathbf{d}_{\theta_d}(\omega) \mathbf{d}_{\theta_d}^H(\omega) \right] \\ &\leq M \text{tr} \left[\mathbf{\Gamma}_v^{-1}(\omega) \right], \end{aligned} \quad (15)$$

where $\text{tr}[\cdot]$ denotes the trace of a square matrix. We observe how the gain is upper bounded [as long as $\mathbf{\Gamma}_v(\omega)$ is nonsingular] and depends on the number of microphones as well as on the nature of the noise.

IV. THE MVDR BEAMFORMER

The MVDR beamformer is derived by minimizing the variance of the residual noise at the beamformer's output with the constraint that signal from the desired look direction is passed through without any distortion. Mathematically, this problem can be written as

$$\mathbf{h}_{\theta_d}(\omega) = \arg \min_{\mathbf{h}(\omega)} \phi_{V_{\text{rn}}}(\omega) \text{ subject to } \mathbf{h}^H(\omega) \mathbf{d}_{\theta_d}(\omega) = 1, \quad (16)$$

where

$$\phi_{V_{\text{rn}}}(\omega) \triangleq E \left[|\mathbf{h}^H(\omega) \mathbf{v}(\omega)|^2 \right] \quad (17)$$

is the variance of the residual noise at the beamformer's output. Using a Lagrange multiplier to adjoin the constraint to the objective function, then differentiating with respect to $\mathbf{h}(\omega)$, and equating the result to zero, we deduce the solution to (16) as [23]

$$\mathbf{h}_{\theta_d}(\omega) = \frac{\mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega)}{\mathbf{d}_{\theta_d}^H(\omega) \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega)}. \quad (18)$$

It is seen that the MVDR beamformer is a function of two terms. One is the steering vector corresponding to the desired signal, which is in turn a function of the incidence angle of the desired signal, and the other is the pseudo-coherence matrix of the observed noise vector.

V. THE SNR GAIN AND BEAMPATTERN OF THE MVDR BEAMFORMER

Substituting (18) into (13) and using the distortionless constraint, one can readily deduce the SNR gain of the MVDR beamformer, i.e.,

$$\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] = \mathbf{d}_{\theta_d}^H(\omega) \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega). \quad (19)$$

This gain has the following property.

Property: The SNR gain of the MVDR beamformer satisfies the following inequalities:

$$1 \leq \mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] \leq M \text{tr} \left[\mathbf{\Gamma}_v^{-1}(\omega) \right], \quad \forall \omega, \theta_d. \quad (20)$$

Proof: $\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] \leq M \text{tr}[\mathbf{\Gamma}_v^{-1}(\omega)]$ is true from (15). Using (14), we can write

$$\mathbf{d}_{\theta_d}^H(\omega) \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega) \geq \frac{|\mathbf{i}^T \mathbf{d}_{\theta_d}(\omega)|^2}{\mathbf{i}^T \mathbf{\Gamma}_v(\omega) \mathbf{i}}, \quad (21)$$

where \mathbf{i} is the first column of the $M \times M$ identity matrix, \mathbf{I}_M . One can check that $\mathbf{i}^T \mathbf{d}_{\theta_d}(\omega) = 1$ and $\mathbf{i}^T \mathbf{\Gamma}_v(\omega) \mathbf{i} = 1$. It follows immediately that $\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] \geq 1$. This completes the proof.

The above property means that the MVDR beamformer can always improve the SNR as compared to the input SNR at the reference sensor;¹ however, the SNR gain is upper bounded, where the upper bound depends on the number of sensors and the statistics of the spatially received noise.

The beampattern of the MVDR beamformer is

$$\mathcal{B}[\mathbf{h}_{\theta_d}(\omega), \theta] = \frac{\mathbf{d}_{\theta}^H(\omega) \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega)}{\mathbf{d}_{\theta_d}^H(\omega) \mathbf{\Gamma}_v^{-1}(\omega) \mathbf{d}_{\theta_d}(\omega)}. \quad (22)$$

It is seen that both the SNR gain and beampattern are a function of the incidence angle of the desired source and the pseudo-coherence matrix of the noise. In what follows, we study these two performance metrics in four different noise scenarios: spatially white noise, diffuse noise, diffuse-plus-white noise, and point-source-plus-white noise.

A. Spatially White Noise

If the noise is spatially white, we have $\mathbf{\Gamma}_v(\omega) = \mathbf{\Gamma}_{\text{wn}}(\omega) = \mathbf{I}_M$, and it can be readily verified that

$$\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] = M, \quad (23)$$

which is a constant and does not depend on the frequency or position of the desired source.

¹The MVDR beamformer can always improve the SNR from a theoretical perspective. However, one may see noise amplification with this beamformer at very low frequencies in practice; this is mainly due to the ill-conditioning or rank-deficiency of the estimated noise covariance matrix.

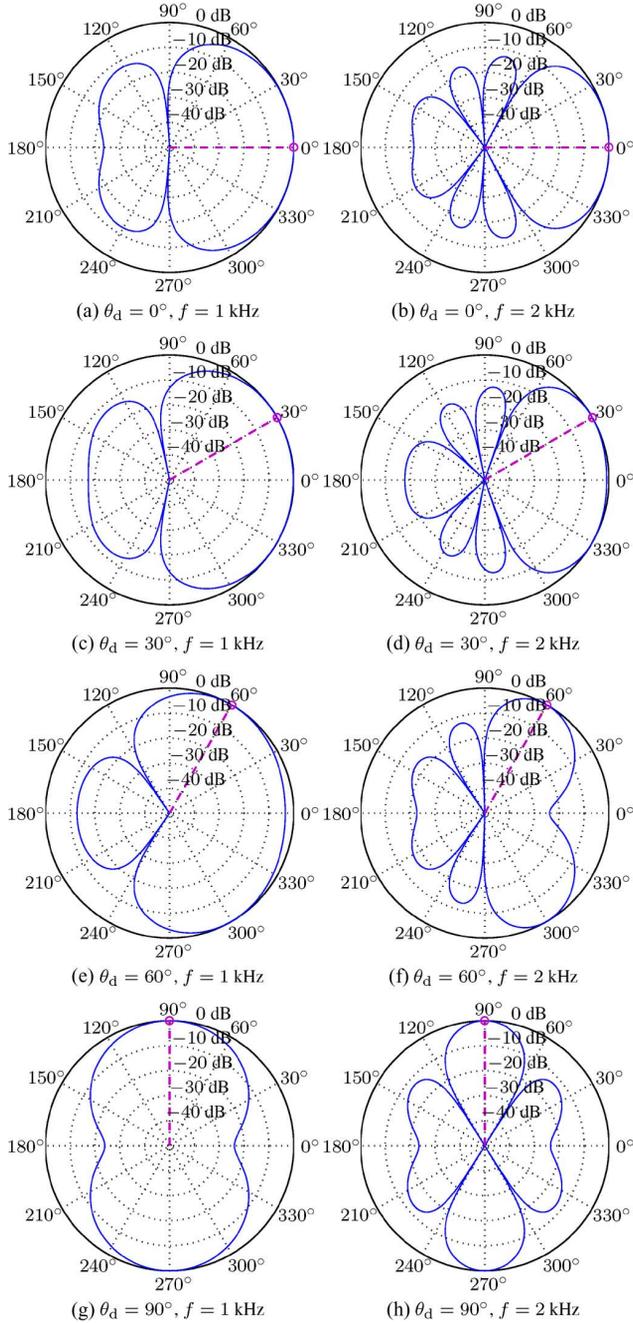


Fig. 2. Beampatterns of the MVDR beamformer with a linear array ($\delta = 4$ cm and $M = 8$) in spatially white noise (where the incidence angle of the desired source is illustrated by the dashed purple line).

The beampattern in this case is

$$\begin{aligned} \mathcal{B}[\mathbf{h}_{\theta_d}(\omega), \theta] &= \frac{\mathbf{d}_{\theta}^H(\omega)\mathbf{d}_{\theta_d}(\omega)}{M} \\ &= \frac{\sin\left[\frac{M}{2}\omega\tau_0(\cos\theta - \cos\theta_d)\right]}{M \sin\left[\frac{1}{2}\omega\tau_0(\cos\theta - \cos\theta_d)\right]}. \end{aligned} \quad (24)$$

Fig. 2 plots the beampattern of the MVDR beamformer according to (24) for different frequencies and incidence angles. Unlike the SNR gain, which is a constant, one can see from Fig. 2 that the beampattern varies significantly with the frequency and the incidence angle.

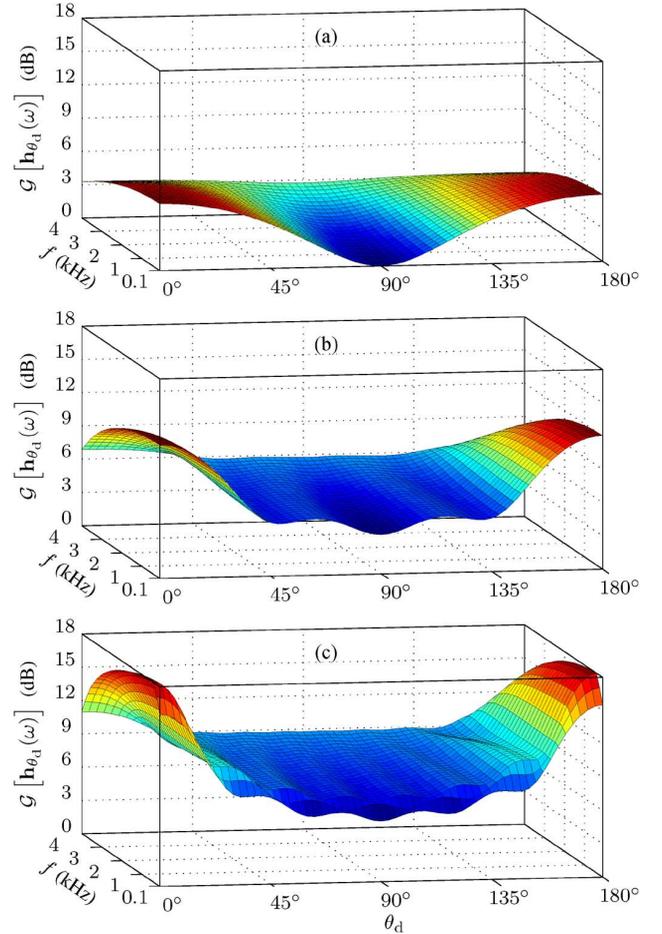


Fig. 3. SNR gain of the MVDR beamformer with a linear array ($\delta = 4$ cm) in diffuse noise: (a) $M = 2$, (b) $M = 4$, and (c) $M = 8$.

B. Diffuse Noise

In reverberant acoustic environments, the noise may have an energy flow of equal probability in all directions, leading to a diffuse noise field [26]–[28]. In this scenario, we have $\mathbf{\Gamma}_v(\omega) = \mathbf{\Gamma}_{\text{dn}}(\omega)$, with

$$[\mathbf{\Gamma}_{\text{dn}}(\omega)]_{ij} = \text{sinc}[\omega\tau_0(j-i)] = \frac{\sin[\omega\tau_0(j-i)]}{\omega\tau_0(j-i)}, \quad (25)$$

where $[\mathbf{\Gamma}_{\text{dn}}(\omega)]_{ij}$ is the (i, j) th element of the matrix $\mathbf{\Gamma}_{\text{dn}}(\omega)$. There are two extreme cases: 1) if $\omega\tau_0$ is very large, e.g., high frequencies or large spacing, the noise signals observed by two sensors tend to be uncorrelated, and then the diffuse noise field is close to the spatially white noise field; 2) if $\omega\tau_0$ is very small, e.g., low frequencies or small spacing, the noise signals observed by two sensors tend to be coherent.

In a two-element array case, we can derive that the SNR gain is

$$\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] = \frac{2[1 - \text{sinc}(\omega\tau_0)\cos(\omega\tau_0\cos\theta_d)]}{1 - \text{sinc}^2(\omega\tau_0)}. \quad (26)$$

It is easy to check that, if $\omega\tau_0 \in [0, \pi]$, the gain in SNR reaches its maximum at $\theta_d = 0^\circ$ and $\theta_d = 180^\circ$, and has a minimum at $\theta_d = 90^\circ$. However, for $M > 2$, it is not easy to find an explicit form of the SNR gain as it is difficult to write $\mathbf{\Gamma}_{\text{dn}}^{-1}(\omega)$ into an analytic form. Fig. 3 plots the SNR gain of the MVDR

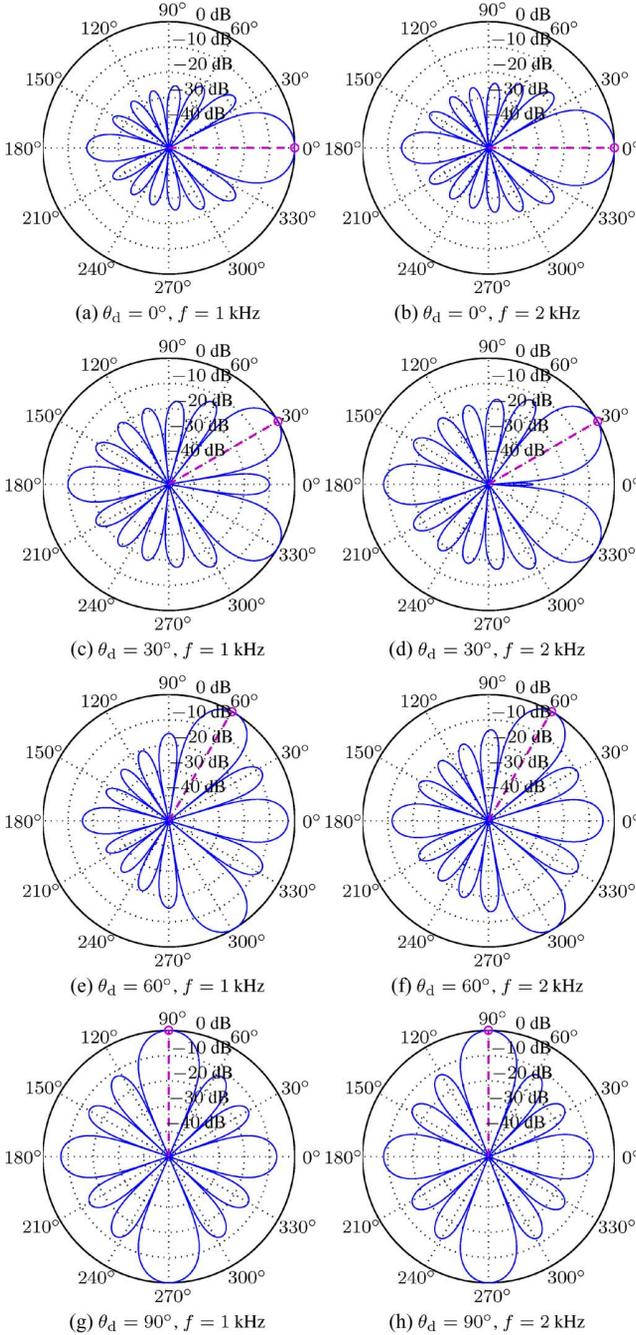


Fig. 4. Beampatterns of the MVDR beamformer with a linear array ($\delta = 4 \text{ cm}$ and $M = 8$) in diffuse noise (where the incidence angle of the desired source is illustrated by the dashed purple line).

beamformer as a function of the desired incidence angle for a linear array with different numbers of microphones. It is seen that the SNR gain increases with the number of microphones. If the number of microphones is fixed, the performance of the MVDR beamformer varies as a function of the desired incidence angle, but the optimal performance always occurs in the endfire directions regardless of the frequency and the number of sensors in the array.

Fig. 4 plots the beampatterns of an 8-element array for different frequencies at four different incidence angles: 0° , 30° , 60° , and 90° . Interestingly, the beampatterns provide another

viewpoint to explain why the MVDR beamformer has a better SNR gain in the endfire directions. First, the beampattern has only one mainlobe in the endfire direction while it is symmetrical and has two mainlobes when the source is from 30° , 60° , and 90° . Although the mainlobe is narrower at $\theta_d = 30^\circ$, 60° , 90° than at $\theta_d = 0^\circ$, the beamformer rejects less noise because it has two mainlobes. Secondly, the height of the sidelobes increases as θ_d increases from 0° to 90° , which again leads to less noise reduction.

C. Diffuse-Plus-White Noise

In the case of diffuse-plus-white noise, the noise pseudo-coherence matrix can be written as

$$\mathbf{\Gamma}_v(\omega) = \mathbf{\Gamma}_{\text{dwn}}(\omega) = (1 - \alpha_{\text{dn}})\mathbf{I}_M + \alpha_{\text{dn}}\mathbf{\Gamma}_{\text{dn}}(\omega), \quad (27)$$

where α_{dn} ($0 \leq \alpha_{\text{dn}} \leq 1$) is a constant that specifies the level of the diffuse noise relative to the spatially white noise, and $\mathbf{\Gamma}_{\text{dn}}(\omega)$ is defined in (25). Similar to the previous case, if $\omega\tau_0$ is very large, the noise signals observed by different sensors tend to be uncorrelated, just like the spatially white noise field.

For the special case of a two-element array, the SNR gain is

$$\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] = \frac{2[1 - \alpha_{\text{dn}}\text{sinc}(\omega\tau_0)\cos(\omega\tau_0\cos\theta_d)]}{1 - \alpha_{\text{dn}}^2\text{sinc}^2(\omega\tau_0)}. \quad (28)$$

Again, if $\omega\tau_0 \in [0, \pi]$, the SNR gain reaches its maximum and minimum in the endfire and broadside directions, respectively. It is difficult to write the SNR gain into an analytic form for $M > 2$. Fig. 5 plots the SNR gain of the MVDR beamformer as a function of the incidence angle and frequency for three cases: 2, 4, and 8 microphones. In all cases, the gain in SNR reaches its maximum in the endfire directions.

Fig. 6 shows the beampatterns of an 8-element array for two different frequencies at four different incidence angles in a diffuse-plus-white noise environment with $\alpha_{\text{dn}} = 0.9$. Again, the beampatterns show why the MVDR beamformer has a better SNR gain in the endfire directions.

One important thing that should be pointed out is that the MVDR beamformer with a linear array has a limited capability of steering. As seen in Fig. 6(c) and (d), the mainlobe points to the endfire direction even though the target direction is $\theta_d = 30^\circ$.

D. Point-Source-Plus-White Noise

In many application scenarios, there may be competing sources. In this subsection, we consider the case where there is a point noise source in addition to the spatially white noise. Assuming that the incidence angle of the point noise source is θ_n , the corresponding pseudo-coherence matrix can be written as

$$\mathbf{\Gamma}_{\text{psn}}(\omega) = \mathbf{d}_{\theta_n}(\omega)\mathbf{d}_{\theta_n}^H(\omega), \quad (29)$$

where $\mathbf{d}_{\theta_n}(\omega)$ is the steering vector of the point noise source, which is defined in a similar way to $\mathbf{d}_{\theta}(\omega)$. Then, the pseudo-coherence matrix of the point-source-plus-white noise is

$$\mathbf{\Gamma}_{\text{pswn}}(\omega) = (1 - \alpha_{\text{psn}})\mathbf{I}_M + \alpha_{\text{psn}}\mathbf{\Gamma}_{\text{psn}}(\omega). \quad (30)$$

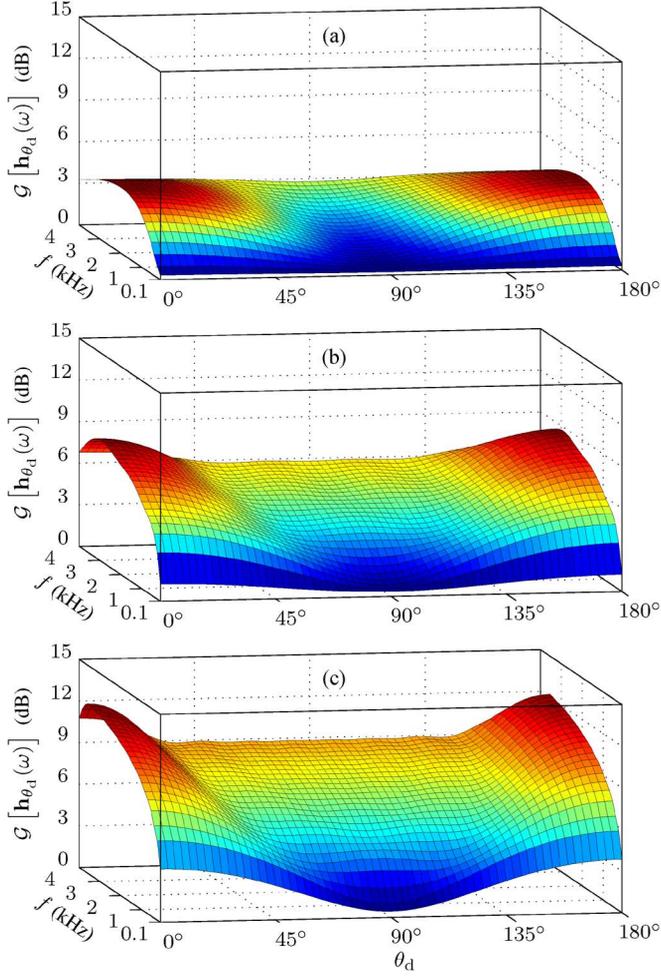


Fig. 5. SNR gain of the MVDR beamformer with a linear array ($\delta = 4$ cm) in diffuse-plus-white noise with $\alpha_{dn} = 0.9$: (a) $M = 2$, (b) $M = 4$, and (c) $M = 8$.

where α_{psn} ($0 \leq \alpha_{psn} < 1$) is a parameter that controls the level of the point source noise relative to that of the spatially white noise. By utilizing the Woodbury's identity, we can express the inverse of the pseudo-coherence matrix as

$$\mathbf{\Gamma}_{pswn}^{-1}(\omega) = \frac{1}{1 - \alpha_{psn}} \left[\mathbf{I}_M - \frac{\mathbf{d}_{\theta_n}(\omega)\mathbf{d}_{\theta_n}^H(\omega)}{(1 - \alpha_{psn})/\alpha_{psn} + M} \right]. \quad (31)$$

Substituting (31) into (19), we can derive the gain in SNR:

$$\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] = \frac{1}{1 - \alpha_{psn}} \left[M - \frac{|\mathbf{d}_{\theta_n}^H(\omega)\mathbf{d}_{\theta_d}(\omega)|^2}{(1 - \alpha_{psn})/\alpha_{psn} + M} \right]. \quad (32)$$

The minimum of the gain in (32) occurs when the desired signal and the point source noise come from the same direction, i.e., $\theta_n = \theta_d$. In this case, the SNR gain is

$$\mathcal{G}[\mathbf{h}_{\theta_d}(\omega)] = \frac{M}{1 + (M - 1)\alpha_{psn}}. \quad (33)$$

Figs. 7 and 8 plot the SNR gain in point-source-plus-white noise environments. In Fig. 7 the point noise source is at $\theta_n = 90^\circ$ while in Fig. 8 the point noise source is at $\theta_n = 135^\circ$. It can be clearly seen that when the point-source noise is present,

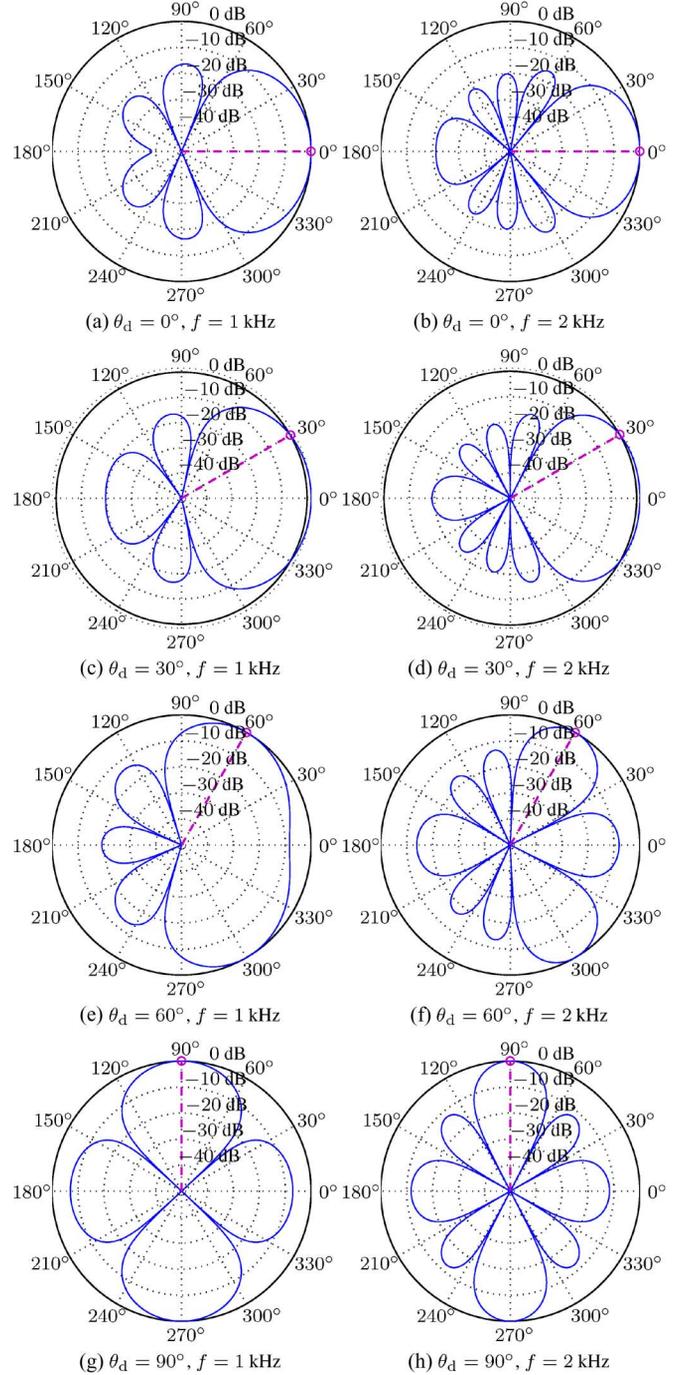


Fig. 6. Beam patterns of the MVDR beamformer with a linear array ($\delta = 4$ cm and $M = 8$) in diffuse-plus-white noise with $\alpha_{dn} = 0.9$ (where the incidence angle of the desired source is illustrated by the dashed purple line).

the SNR gain depends on both the angular separation between the point-noise and desired sources as well as the number of sensors.

Another important factor that affects the performance of the MVDR beamformer in point-source-plus-white noise environments is the level of the point-source noise. Fig. 9 plots the SNR gain as a function of the parameter α_{psn} . If the point-noise and desired sources are not incident from the same direction, the SNR gain increases monotonously with α_{psn} . It is possible for the MVDR filter to completely reject the noise when $\alpha_{psn} = 1$. However, this extreme case is beyond the scope of this paper

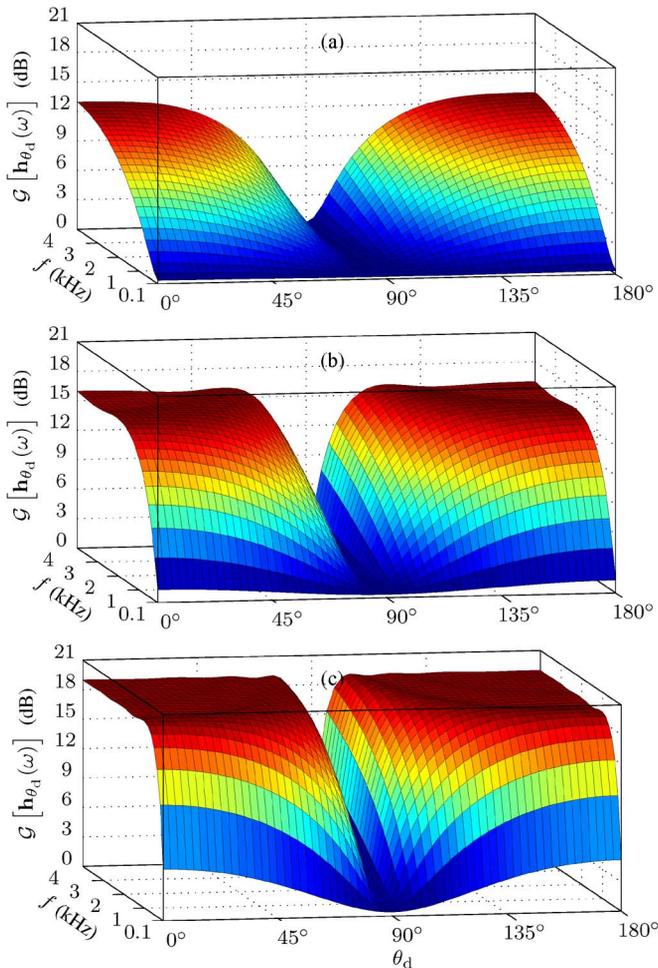


Fig. 7. SNR gain of the MVDR beamformer with a linear array ($\delta = 4$ cm) in point-source-plus-white noise with $\alpha_{\text{psn}} = 0.9$ and $\theta_n = 90^\circ$: (a) $M = 2$, (b) $M = 4$, and (c) $M = 8$.

as the noise pseudo-coherence matrix, $\Gamma_v(\omega)$, is no longer full rank.

Fig. 10 shows the beampatterns of an 8-element array in point-source-plus-white noise environments. These beampatterns illustrate how the point source noise is reduced. The MVDR beamformer puts a null at θ_n to reduce the point source noise as long as there is a good angular separation between the noise and desired sources.

VI. PERFORMANCE EVALUATION IN REVERBERANT ACOUSTIC ENVIRONMENTS

In real acoustic environments, the signals picked up by a microphone consists of not only the desired source signal and noise but also reverberation. In this situation, the microphone array outputs are no longer in the simple form of (1), but in a more complicated form as [21]

$$\begin{aligned} y_m(t) &= g_m(t) * s(t) + v_m(t) \\ &= x_m(t) + v_m(t), \quad m = 1, 2, \dots, M, \end{aligned} \quad (34)$$

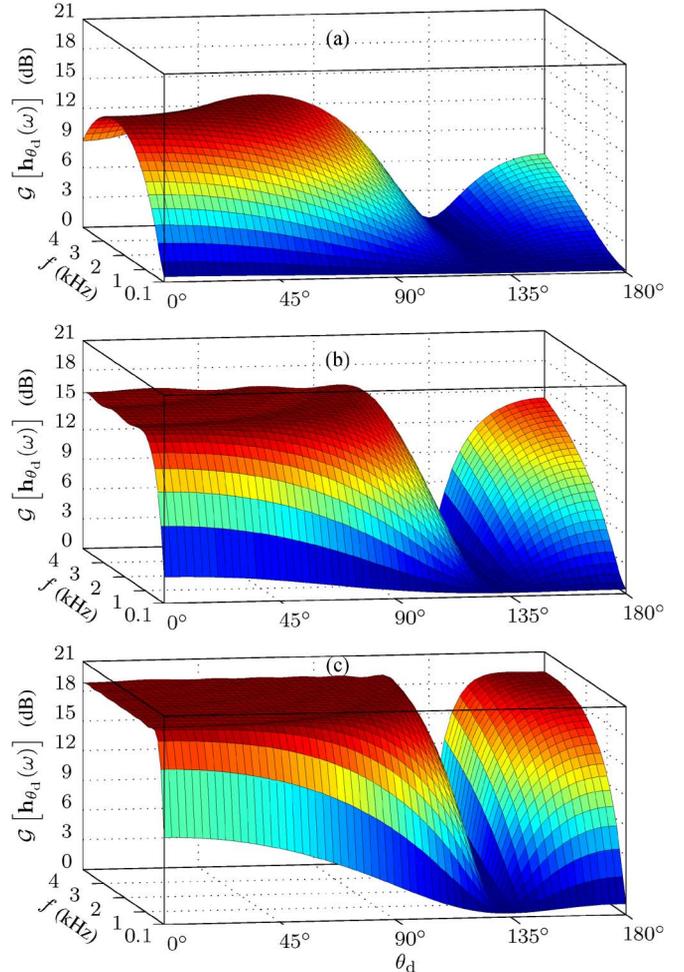


Fig. 8. SNR gain of the MVDR beamformer with a linear array ($\delta = 4$ cm) in point-source-plus-white noise with $\alpha_{\text{psn}} = 0.9$ and $\theta_n = 135^\circ$: (a) $M = 2$, (b) $M = 4$, and (c) $M = 8$.

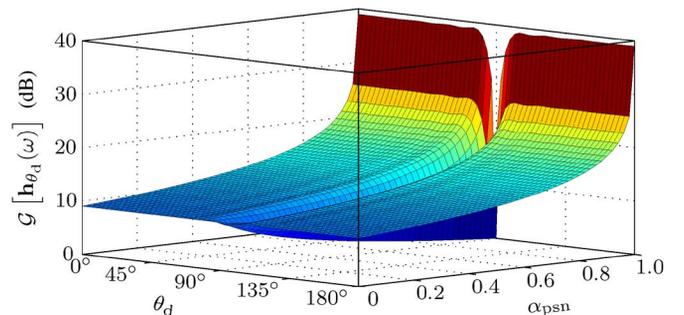


Fig. 9. SNR gain of the MVDR beamformer with a linear array ($\delta = 4$ cm and $M = 8$) as a function of α_{psn} in point-source-plus-white noise, where $\theta_n = 90^\circ$ and $f = 4$ kHz.

where $g_m(t)$ is the impulse response from the desired source, $s(t)$, to the m th microphone and $*$ denotes linear convolution. The corresponding frequency-domain counterpart is written as

$$\begin{aligned} Y_m(\omega) &= G_m(\omega)S(\omega) + V_m(\omega) \\ &= X_m(\omega) + V_m(\omega), \quad m = 1, 2, \dots, M, \end{aligned} \quad (35)$$

where $G_m(\omega)$ and $S(\omega)$ are the Fourier transforms of $g_m(t)$ and $s(t)$, respectively. In reverberation conditions, the steering vector of the source signal, i.e., $\mathbf{d}_{\theta_d}(\omega)$ in (4), is no longer a

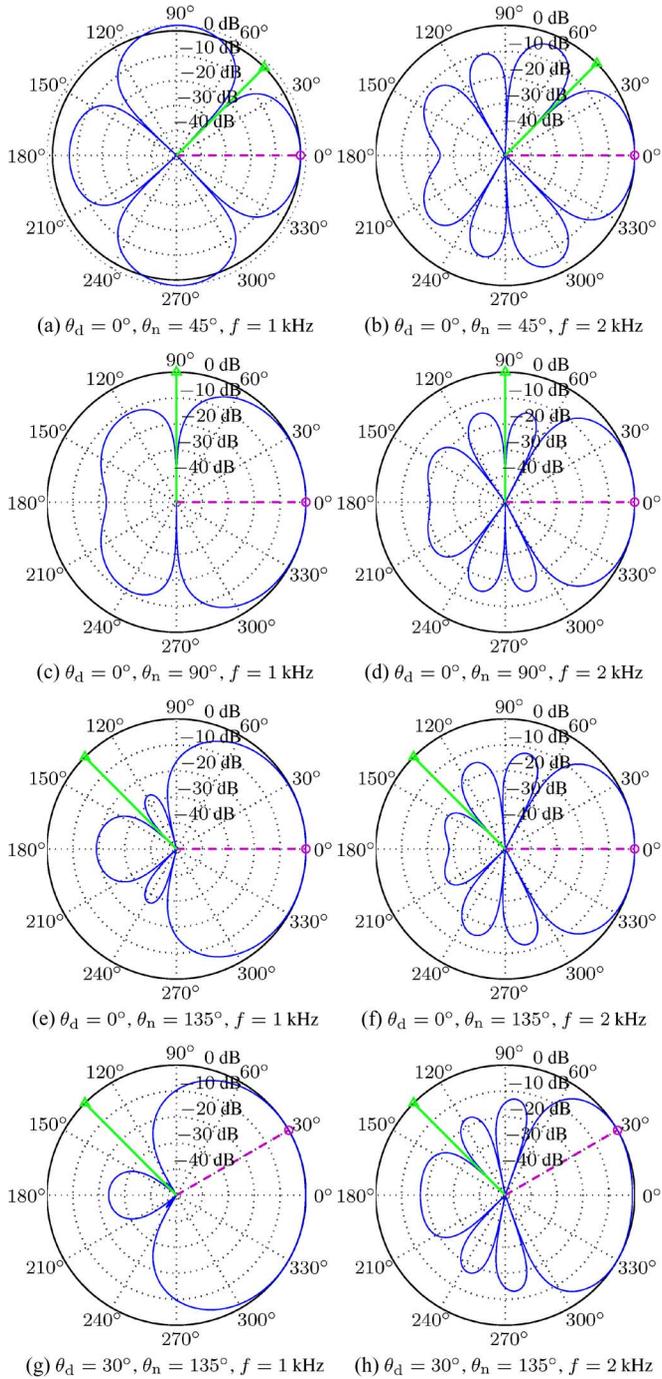


Fig. 10. Beam patterns of the MVDR beamformer with a linear array ($\delta = 4 \text{ cm}$ and $M = 8$) in point-source-plus-white noise with $\alpha_{\text{psn}} = 0.9$ (the incidence angle of the desired source is illustrated by the dashed purple line and the incidence angle of the point-noise source is illustrated by the solid bright green line).

simple function of the incidence angle θ_d . As a result, the performance of the MVDR beamformer may degrade significantly, depending on the degree of reverberation. In this section, we investigate the performance of the MVDR beamformer in reverberant environments through simulations.

A. Simulation Setup

A diagram of the floor layout of the simulation setup is shown in Fig. 11. A linear microphone array with a total of 8 micro-

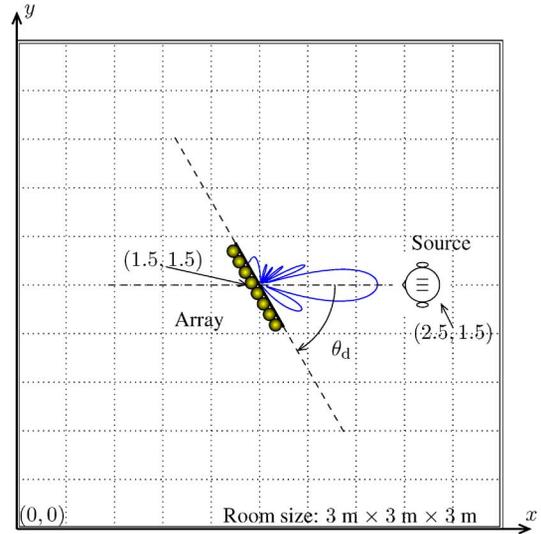


Fig. 11. Floor layout of the simulation setup.

phones is placed in a reverberant room of size $3 \text{ m} \times 3 \text{ m} \times 3 \text{ m}$. Both the microphone array and the speech source are on the horizontal plane at $z = 1.5 \text{ m}$. For ease of exposition, positions in the floor plan are designated by (x, y) coordinates with reference to the southwest corner. The center of the linear array is located at $(1.5, 1.5)$. The distance between two neighboring sensors is 4 cm . A desired speech source is placed at $(2.5, 1.5)$. For simplicity, we assume that the reflection coefficients β_i ($i = 1, 2, \dots, 6$) of all the six walls are identical, i.e., $\beta_1 = \beta_2 = \dots = \beta_6 = \beta$, which varies between 0 to 1. To study the impact of the source incidence angle on the performance of the MVDR beamformer, we fix the source position, but rotate the array clockwise with respect to the array center so that the source incidence angle changes from 0° to 180° . We assume that the incidence angle θ_d is known to the MVDR beamformer and put our focus on studying how the performance of the MVDR beamformer would change with respect to θ_d .

The room impulse responses from the source position to the microphone sensors are generated with the well-known image-model method [22], [29]. Fig. 12 shows an example of the generated impulse responses with $\beta = 0.8$ and the corresponding reverberation time $T_{60} \approx 215 \text{ ms}$. The variable T_{60} is defined as the time for the sound to die away to a level 60 dB below its original level and is measured by the Schroeder's method [29] using the reverse-time integrated impulse response.

The desired source is a speech signal recorded in a quiet room with a sampling rate of 8 kHz . The length of this signal is 25 s . The microphone array outputs are generated by convolving the source signal with the generated impulse responses. Noise is then added to the convolution result to control the input SNR. We consider four different types of noise, i.e., spatially white noise, diffuse noise, diffuse-plus-white noise, and point-source-plus-white noise. The spatially white Gaussian noise is generated using the Matlab `randn` function. The diffuse noise is generated by the method presented in [31], which sums 100 point sources (each source is a white noise) uniformly distributed on the surface of a sphere around the array. The diffuse-plus-white noise is generated by scaling (to control the level of the diffuse

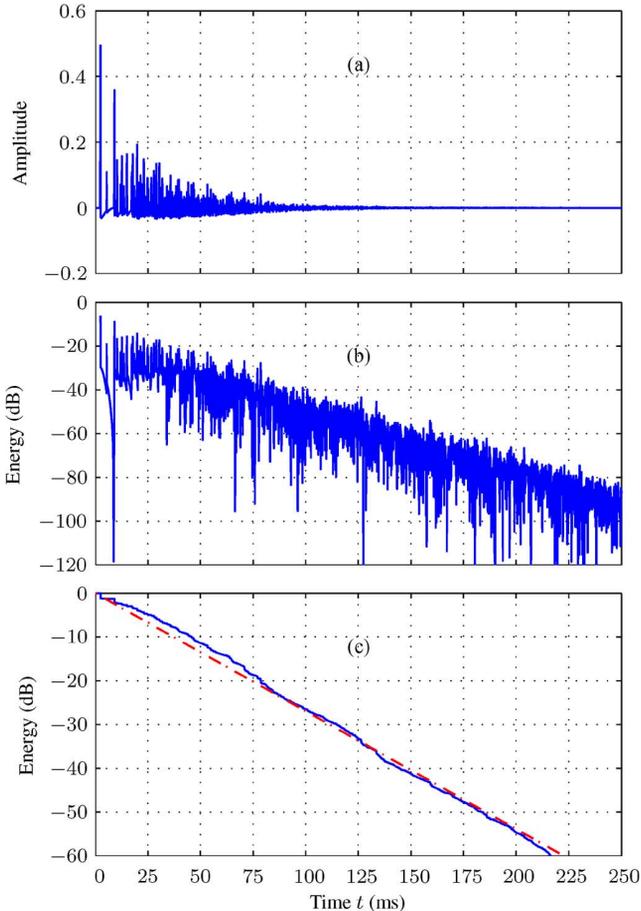


Fig. 12. Acoustic impulse response from the desired source to the first microphone: (a) impulse response generated by the image-model method, (b) magnitude square of the impulse response, and (c) backward integration of the squared impulse response and a linear fitting curve. All the reflection coefficients are 0.8, $\theta_d = 0^\circ$, and the sampling rate is 8 kHz.

noise relative to the spatially white noise) and adding the spatially white noise and diffuse noise together. The point-noise source is also a white noise signal. The point-source noise observed at each microphone is generated by convolving the point-noise signal with the impulse response from the noise source to the microphone.

Note that the direct path positions are integral multiples of the sampling period using the image-model method (while fractional delays are possible, we consider only integral delays in this paper for simplicity). Furthermore, the source is not completely in the far field with the given room size and array configuration. Therefore, the relative time delay between the m th microphone and the reference sensor (i.e., the time difference between two direct paths) may not strictly satisfy the relation $\tau_m = (m-1)\tau_0 \cos \theta_d$ ($m = 1, 2, \dots, M$). In our simulations, we slightly time shift the microphone signals after the convolution between the source signal and impulse responses so that $\tau_m = (m-1)\tau_0 \cos \theta_d$ is satisfied. One should note, however, that this compensation does not change the reverberation conditions, and therefore, will not affect the observations and conclusions achieved in the following simulations.

In Sections III and V, the SNR gain is defined and evaluated on a narrowband basis. Now, in this section, we start to evaluate the full-band SNR gain in the time domain, which is defined as

$$\mathcal{G} \triangleq \frac{\text{oSNR}}{\text{iSNR}}, \quad (36)$$

where the full-band input SNR is defined as

$$\text{iSNR} \triangleq \frac{E[x_1^2(t)]}{E[v_1^2(t)]} \quad (37)$$

and the full-band output SNR is given by

$$\text{oSNR} \triangleq \frac{E[x_{\text{id}}^2(t)]}{E[v_{\text{rn}}^2(t)]}, \quad (38)$$

where $x_{\text{id}}(t)$ and $v_{\text{rn}}(t)$ are the time-domain filtered desired signal and residual noise reconstructed from $\mathbf{h}_{\theta_d}^H(\omega)\mathbf{x}(\omega)$ and $\mathbf{h}_{\theta_d}^H(\omega)\mathbf{v}(\omega)$, respectively.

B. MVDR Beamformer Implementation

The MVDR beamformer is implemented with the overlap-add technique, which is widely used in speech enhancement and noise reduction [34]. We first divide the microphones' outputs into overlap frames with a frame width of 128 (16 ms with a sampling rate of 8 kHz) and an overlapping factor of 75%. Each frame is multiplied with a Kaiser window and transformed into the short-time Fourier transform (STFT) domain. In each subband, an MVDR beamformer is designed and applied to the noisy signals. Finally, the time-domain beamforming output is constructed by transforming the MVDR filtered signal back to the time domain.

The implementation of the MVDR beamformer in each subband would require the estimation of the noise pseudo-coherence matrix, which is now denoted as $\mathbf{\Gamma}_v(k, n)$, where k and n are the frequency bin and time frame, respectively. In this paper, we directly compute the noise covariance matrix, $\hat{\mathbf{\Phi}}_v(k, n)$, using a short-time average (so the nonstationarity of speech and noise can be taken into account) with the most recent 100 frames. Then, an estimate of the pseudo-coherence matrix $\hat{\mathbf{\Gamma}}_v(k, n)$ is computed according to (6).

The pseudo-coherence matrix $\hat{\mathbf{\Gamma}}_v(k, n)$ can be ill conditioned at low frequencies in diffuse noise, which may lead to numerical instability in implementation of the MVDR beamformer. To circumvent this issue and achieve a robust computation of the inverse of $\hat{\mathbf{\Gamma}}_v(k, n)$, the pseudo inverse technique with eigenvalue decomposition is used.² First, the Hermitian matrix $\hat{\mathbf{\Gamma}}_v(k, n)$ is decomposed into

$$\hat{\mathbf{\Gamma}}_v = \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H = \mathbf{Q} \begin{bmatrix} \mathbf{\Lambda}_P & \mathbf{0} \\ \mathbf{0} & \mathbf{\Lambda}_{M-P} \end{bmatrix} \mathbf{Q}^H, \quad (39)$$

where $\mathbf{\Lambda} = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_M)$ is the diagonal eigenvalue matrix with $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_M$, \mathbf{Q} is the corresponding unitary matrix consisting of all the eigenvectors, $\mathbf{\Lambda}_P$ is a sub

²In practical implementations, the diagonal loading [32] and white-noise-gain constraint [33] methods can also be used to deal with the matrix ill-conditioning issue. But these approaches are not appropriate to study the diffuse only noise case.

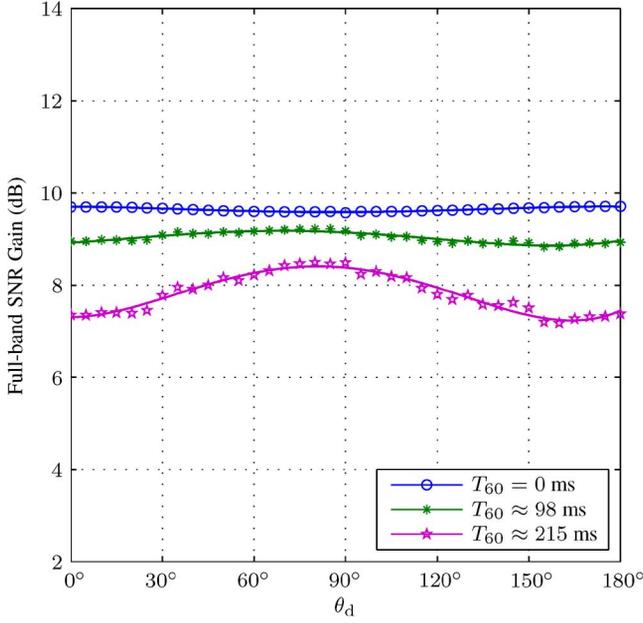


Fig. 13. Full-band SNR gain in spatially white noise with $T_{60} = 0$ ($\beta = 0$), $T_{60} \approx 98$ ms ($\beta = 0.6$) and $T_{60} \approx 215$ ms ($\beta = 0.8$). The fitting curve is a fourth-order polynomial.

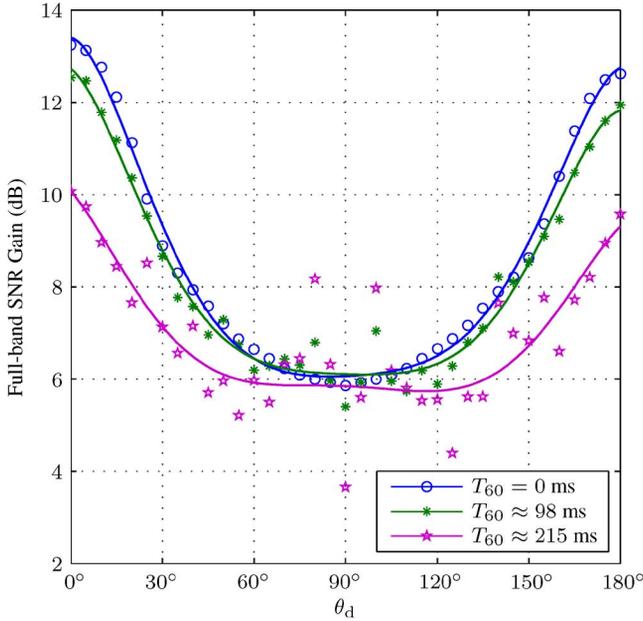


Fig. 14. Full-band SNR gain in diffuse noise with $T_{60} = 0$ ($\beta = 0$), $T_{60} \approx 98$ ms ($\beta = 0.6$) and $T_{60} \approx 215$ ms ($\beta = 0.8$). The fitting curve is a sixth-order polynomial.

diagonal matrix consisting of the $P (\leq M)$ largest eigenvalues of $\hat{\Gamma}_{\mathbf{v}}$, and $\mathbf{\Lambda}_{M-P}$ is a sub diagonal matrix consisting of the $M - P$ smallest eigenvalues of $\hat{\Gamma}_{\mathbf{v}}$. Then, the inverse of $\hat{\Gamma}_{\mathbf{v}}$ is computed as

$$\hat{\Gamma}_{\mathbf{v}}^{-1} = \mathbf{Q} \begin{bmatrix} \mathbf{\Lambda}_P^{-1} & \mathbf{0} \\ \mathbf{0} & \mathbf{0}_{M-P} \end{bmatrix} \mathbf{Q}^H. \quad (40)$$

The underlying assumption is that all the $M - P$ very small eigenvalues can be neglected. In implementation, P is time-varying and is chosen according to $\lambda_P \geq \xi$ and $\lambda_{P+1} < \xi$, where ξ is a threshold and set to $10^{-4} \times \lambda_1$ in our simulations.

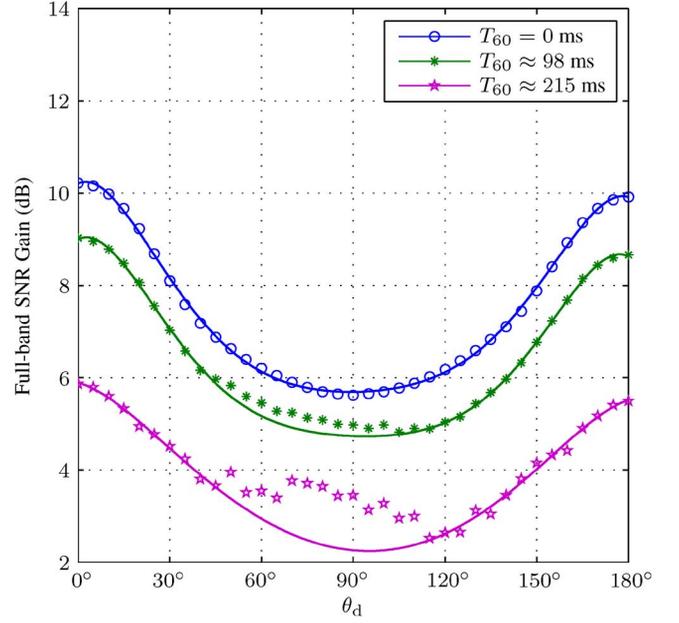


Fig. 15. Full-band SNR gain in diffuse-plus-white noise ($\alpha_{dn} = 0.9$) with $T_{60} = 0$ ($\beta = 0$), $T_{60} \approx 98$ ms ($\beta = 0.6$) and $T_{60} \approx 215$ ms ($\beta = 0.8$). The fitting curve is a fifth-order polynomial.

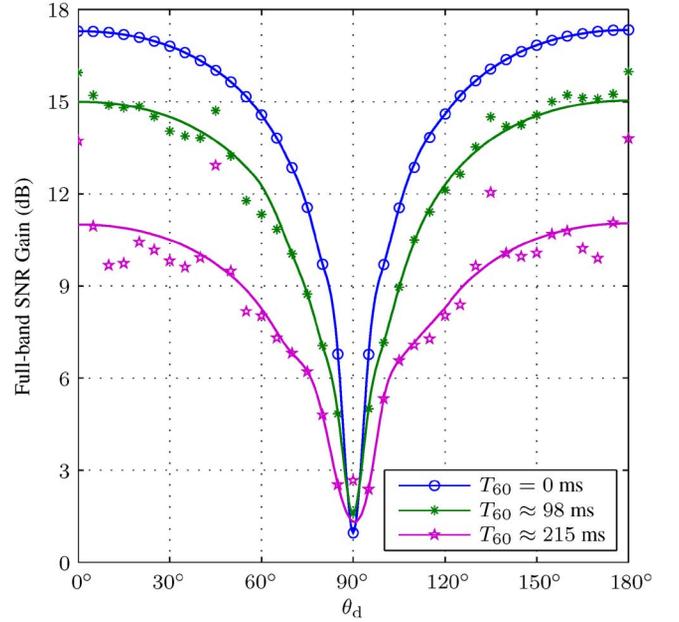


Fig. 16. Full-band SNR gain in point-source-plus-white noise ($\alpha_{pn} = 0.9$) with $T_{60} = 0$ ($\beta = 0$), $T_{60} \approx 98$ ms ($\beta = 0.6$) and $T_{60} \approx 215$ ms ($\beta = 0.8$). The point noise source is incident from 90° .

C. Simulation Results

The SNR gains of the implemented MVDR beamformer in different reverberant and noise environments are plotted, respectively, in Figs. 13, 14, 15, and 16.

• Spatially white noise.

In spatially white noise and when there is no reverberation, i.e., $\beta = 0$ and $T_{60} = 0$, it is seen from Fig. 13 that the full-band SNR gain of the MVDR beamformer is a constant and does not change with the signal incidence angle, which corroborates with the theoretical analysis in Section V. The theoretical SNR gain in this case should be 9 dB. The simulation result is slightly higher than 9 dB.

This minor difference in SNR gain could be due to the use of the short-time average to compute the noise correlation matrix.

When reverberation is present ($T_{60} \approx 98$ ms and $T_{60} \approx 215$ ms), there is some degradation in SNR gain, but the SNR gain does not change much with respect to the signal incidence angle, which still agrees well with the theoretical analysis in Section V for the case with no reverberation.

- **Diffuse noise.**

The simulation results in diffuse noise and reverberation conditions are sketched in Fig. 14. It is clearly seen that the MVDR beamformer achieves the maximum SNR gain in the endfire directions regardless of the degree of reverberation.

Similar to the previous simulation in spatially white noise, one can see that reverberation can cause some degradation in SNR gain. Besides the reverberation time, the structure of reverberation may also affect the SNR gain, as seen from Fig. 14 that the measured SNR gain may deviate dramatically from the fitting curve in some directions.

- **Diffuse-plus-white noise.**

The full-band SNR gain in a diffuse-plus-white noise and reverberant environments are shown in Fig. 15. It is seen that the presence of reverberation can cause degradation in SNR gain. However, the MVDR beamformer always obtains its maximum SNR gain in the endfire directions in both the anechoic and reverberant conditions.

- **Point-source-plus-white noise.**

In this simulation, a point-noise source is incident to the array from $\theta_n = 90^\circ$ and the constant that controls the level of the point source noise relative to the spatially white noise is $\alpha_{psn} = 0.9$. The full-band SNR gain as a function of the desired signal incidence angle in this simulation is shown in Fig. 16. One can see that there is not much noise reduction when $\theta_d = 90^\circ$. This is due to the fact the point-noise source is also in 90° . But significant SNR improvement is observed in the endfire directions, which, again, corroborates the theoretical study in Section V. Once again, the MVDR beamformer with a linear array achieves its optimal performance in the endfire directions as long as the point-noise source is away from these directions.

VII. CONCLUSIONS

In this paper, we addressed the problem on how to best configure a microphone array system so that the MVDR beamformer can reach its best performance in terms of signal enhancement and noise and interference reduction. Using a linear microphone array, we investigated its performance in four different noise fields: spatially white noise, diffuse noise, diffuse-plus-white noise, and point-source-plus-white noise. Both the theoretical and simulation results demonstrated that the MVDR beamformer's performance strongly depends on the incidence angle of the desired source. It achieves the optimal SNR gain in the endfire directions in diffuse noise. When there is a point-source noise, the SNR gain depends on the angular separation of the point-noise and desired sources as well as the point-noise level. But as long as the point-noise source is not in the end-

fire direction, the optimal SNR gain still occurs in the endfire direction. If there is only spatially white noise, the SNR gain of the MVDR beamformer only depends on the number of sensors and is not dependent on the source incidence angle. From this study, it is safe to conclude that, given a linear microphone array, one should configure it such that the endfire direction is pointed to the desired source in order for the MVDR beamformer to achieve its best performance.

ACKNOWLEDGMENT

The authors are very grateful to the anonymous reviewers for their careful reading of the draft and providing many constructive comments and suggestions that have helped improve the clarity of this paper.

REFERENCES

- [1] S. A. Schelkunoff, "A mathematical theory of linear arrays," *Bell Syst. Tech. J.*, vol. 22, pp. 80–107, Jan. 1943.
- [2] D. E. Dudgeon, "Fundamentals of digital array processing," *Proc. IEEE*, vol. 65, no. 6, pp. 898–904, Jun. 1977.
- [3] J. L. Flanagan, J. D. Johnson, R. Zahn, and G. W. Elko, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Amer.*, vol. 75, pp. 1508–1518, Nov. 1985.
- [4] *Microphone Arrays: Signal Processing Techniques and Applications*, M. Brandstein and D. B. Ward, Eds. Berlin, Germany: Springer-Verlag, 2001.
- [5] D. B. Ward, R. C. Williamson, and R. A. Kennedy, "Broadband microphone arrays for speech acquisition," *Acoust. Australia*, vol. 26, pp. 17–20, Apr. 1998.
- [6] J. L. Flanagan, D. A. Berkeley, G. W. Elko, J. E. West, and M. M. Sondhi, "Autodirective microphone systems," *Acustica*, vol. 73, pp. 58–71, Feb. 1991.
- [7] W. Kellermann, "A self-steering digital microphone array," in *Proc. IEEE ICASSP*, 1991, vol. 5, pp. 3581–3584.
- [8] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, pp. 1053–1065, Mar. 2007.
- [9] O. L. Frost, III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, pp. 926–935, Aug. 1972.
- [10] C. W. Jim, "A comparison of two LMS constrained optimal array structures," *Proc. IEEE*, vol. 65, no. 12, pp. 1730–1731, Dec. 1977.
- [11] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. AP-30, no. 1, pp. 27–34, Jan. 1982.
- [12] M. M. Sondhi and G. W. Elko, "Adaptive optimization of microphone arrays under a nonlinear constraint," in *Proc. IEEE ICASSP*, 1986, pp. 981–984.
- [13] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.
- [14] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.
- [15] W. Herboldt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds. Berlin, Germany: Springer, 2003, ch. 6, pp. 155–194.
- [16] K. M. Buckley, "Broad-band beamforming and the generalized side-lobe canceller," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 5, pp. 1322–1323, Oct. 1986.
- [17] S. Werner, J. A. Apolinário, and M. L. R. de Campos, "On the equivalence of RLS implementations of LCMV and GSC processors," *IEEE Signal Process. Lett.*, vol. 10, no. 12, pp. 356–359, Dec. 2003.
- [18] B. R. Breed and J. Strauss, "A short proof of the equivalence of LCMV and GSC beamforming," *IEEE Signal Process. Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2004.
- [19] C. Marro, Y. Mahieux, and K. U. Simmer, "Analysis of noise reduction and dereverberation techniques based on microphone arrays with postfiltering," *IEEE Trans. Speech Audio Process.*, vol. 6, no. 3, pp. 240–259, May 1998.

- [20] I. Cohen, "Analysis of two-channel generalized sidelobe canceller (GSC) with post-filtering," *IEEE Trans. Speech Audio Process.*, vol. 11, no. 6, pp. 684–699, Nov. 2003.
- [21] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.
- [22] Y. Huang, J. Benesty, and J. Chen, *Acoustic MIMO Signal Processing*. Berlin, Germany: Springer-Verlag, 2006.
- [23] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.
- [24] H. Cox, R. M. Zeskind, and T. Kooij, "Practical supergain," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 3, pp. 393–398, Jun. 1986.
- [25] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-35, no. 10, pp. 1365–1376, Oct. 1987.
- [26] F. Jacobsen, The diffuse sound field: Statistical considerations concerning the reverberant field in the steady state Acoustics Lab., Tech. Univ. of Denmark, 1979, Rep..
- [27] M. Goulding, "Speech enhancement for mobile telephony [microform]," Canadian theses on microfiche, Thesis (M.A.Sc.), Simon Fraser Univ., Burnaby, BC, Canada, 1989.
- [28] *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Berlin, Germany: Springer-Verlag, 2007.
- [29] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *J. Acoust. Soc. Amer.*, vol. 65, pp. 943–950, Apr. 1979.
- [30] E. A. Lehmann and A. M. Johansson, "Prediction of energy decay in room impulse responses simulated with an image-source model," *J. Acoust. Soc. Amer.*, vol. 124, pp. 269–277, Jul. 2008.
- [31] E. A. P. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," *J. Acoust. Soc. Amer.*, vol. 122, pp. 3464–3470, Dec. 2007.
- [32] J. Li, P. Stoica, and Z. Wang, "On robust Capon beamforming and diagonal loading," *IEEE Trans. Signal Process.*, vol. 51, no. 7, pp. 1702–1715, Jul. 2003.
- [33] S. M. Kogon, "Eigenvectors, diagonal loading and white noise gain constraints for robust adaptive beamforming," in *Proc. 37th Asilomar Conf. Signals, Syst., Comput.*, 2004, vol. 2, pp. 1853–1857.
- [34] J. Benesty, J. Chen, Y. Huang, and I. Cohen, *Noise Reduction in Speech Processing*. Berlin, Germany: Springer-Verlag, 2009.



Chao Pan (SM'13) was born in 1989. He received the Bachelor degree in Electronics and Information Engineering from the Northwestern Polytechnical University (NWPU) in 2011. He is currently a Ph.D. student in Information and Communication Engineering at NWPU. His research interests are in speech enhancement, noise reduction, and microphone array signal processing for hands-free speech communications.



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 acoustic signal processing, adaptive signal processing, speech enhancement, adaptive noise/echo control, microphone array signal processing, signal separation, and speech communication. Dr. Chen is currently an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING, an associate member of the IEEE Signal Processing Society (SPS) Technical Committee (TC) on Audio and Acoustic Signal Processing (AASP), and a member of the editorial advisory board of the Open Signal Processing Journal. He was the Technical Program Co-Chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA) and the Technical Program Chair of IEEE TENCON 2013, and helped organize many other conferences. He co-authored the books *Study and Design of Differential Microphone Arrays* (Springer-Verlag, 2013), *Speech Enhancement in the STFT Domain* (Springer-Verlag, 2011), *Optimal Time-Domain Noise Reduction Filters: A Theoretical Study* (Springer-Verlag, 2011), *Speech Enhancement in the Karhunen-Loève Expansion Domain* (Morgan&Claypool, 2011), *Noise Reduction in Speech Processing* (Springer-Verlag, 2009), *Microphone Array Signal Processing* (Springer-Verlag, 2008), and *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006). He is also a co-editor/co-author of the book *Speech Enhancement* (Berlin, Germany: Springer-Verlag, 2005) and a section co-editor of the reference *Springer Handbook of Speech Processing* (Springer-Verlag, Berlin, 2007).

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 (WASPAA) 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, the Japan Trust International Research Grant from the Japan Key Technology Center in 1998, the Young Author Best Paper Award from the 5th National Conference on Man-Machine Speech Communications in 1998, and the CAS (Chinese Academy of Sciences) President's Award in 1998.



Jacob Benesty was born in 1963. He 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 Telecommunications (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. 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 was the co-chair of the 1999 International Workshop on Acoustic Echo and Noise Control and the general co-chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics. He is the recipient, with Morgan and Sondhi, of the IEEE Signal Processing Society 2001 Best Paper Award. He is the recipient, with Chen, Huang, and Doclo, of the IEEE Signal Processing Society 2008 Best Paper Award. He is also the co-author of a paper for which Huang received the IEEE Signal Processing Society 2002 Young Author Best Paper Award. In 2010, he received the "Gheorghe Cartianu Award" from the Romanian Academy. In 2011, he received the Best Paper Award from the IEEE WASPAA for a paper that he co-authored with Chen.