New Insights Into the MVDR Beamformer in Room Acoustics

E. A. P. Habets, *Member, IEEE*, J. Benesty, *Senior Member, IEEE*, I. Cohen, *Senior Member, IEEE*, S. Gannot, *Senior Member, IEEE*, and J. Dmochowski

***Abstract*—The minimum variance distortionless response (MVDR) beamformer, also known as Capon's beamformer, is widely studied in the area of speech enhancement. The MVDR beamformer can be used for both speech dereverberation and noise reduction. This paper provides new insights into the MVDR beamformer. Specifically, the local and global behavior of the MVDR beamformer is analyzed and novel forms of the MVDR filter are derived and discussed. In earlier works it was observed that there is a tradeoff between the amount of speech dereverberation and noise reduction when the MVDR beamformer is used. Here, the tradeoff between speech dereverberation and noise reduction is analyzed thoroughly. The local and global behavior, as well as the tradeoff, is analyzed for different noise fields such as, for example, a mixture of coherent and non-coherent noise fields, entirely non-coherent noise fields and diffuse noise fields. It is shown that maximum noise reduction is achieved when the MVDR beamformer is used for noise reduction only. The amount of noise reduction that is sacrificed when complete dereverberation is required depends on the direct-to-reverberation ratio of the acoustic impulse response between the source and the reference microphone. The performance evaluation supports the theoretical analysis and demonstrates the tradeoff between speech dereverberation and noise reduction. When desiring both speech dereverberation and noise reduction, the results also demonstrate that the amount of noise reduction that is sacrificed decreases when the number of microphones increases.**

***Index Terms*—Beamforming, microphone arrays, minimum variance distortionless response (MVDR) filter, noise reduction, Pearson correlation coefficient, speech dereverberation, speech enhancement.**

I. Introduction

DISTANT or hands-free audio acquisition is required in many applications such as audio-bridging and teleconferencing. Microphone arrays are often used for the acquisition and consist of sets of microphone sensors that are arranged in specific patterns. The received sensor signals usually consist of a desired sound signal, coherent and non-coherent interferences. The received signals are processed in order to extract the desired sound, or in other words to suppress the interferences. In the last four decades, many algorithms have been proposed to process the received sensor signals [1], [2].

The minimum variance distortionless response (MVDR) beamformer, also known as Capon beamformer [3], minimizes the output power of the beamformer under a single linear constraint on the response of the array towards the desired signal. The idea of combining multiple inputs in a statistically optimum manner under the constraint of no signal distortion can be attributed to Darlington [4]. Several researchers developed beamformers in which additional linear constraints were imposed (e.g., Er and Cantoni [5]). These beamformers are known as linearly constrained minimum variance (LCMV) beamformers, of which the MVDR beamformer is a special case. In [6], Frost proposed an adaptive scheme of the MVDR beamformer, which is based on a constrained least-mean-square (LMS) type adaptation. Kaneda *et al.* [7] proposed a noise reduction system for speech signals, termed AMNOR, which adopts a soft-constraint that controls the tradeoff between speech distortion and noise reduction. To avoid the constrained adaptation of the MVDR beamformer, Griffiths and Jim [8] proposed the generalized sidelobe canceller (GSC) structure, which separates the output power minimization and the application of the constraint. While Griffiths and Jim only considered one constraint (i.e., MVDR beamformer), it was later shown in [9] that the GSC structure can also be used in the case of multiple constraints (i.e., LCMV beamformer). The original GSC structure is based on the assumption that the different sensors receive a delayed version of the desired signal. The GSC structure was re-derived in the frequency-domain, and extended to deal with general acoustic transfer functions (ATFs) by Affes and Grenier [10] and later by Gannot *et al.* [11]. The frequency-domain version in [11], which takes into account the reverberant nature of the enclosure, was termed the transfer-function generalized sidelobe canceller (TF-GSC).

In theory, the LCMV beamformer can achieve perfect dereverberation and noise cancellation when the ATFs between all sources (including interferences) and the microphones are known [12]. Using the MVDR beamformer, we can achieve perfect reverberation cancellation when the ATFs between the desired source and the microphones are known. In the last three decades, various methods have been developed to blindly identify the ATFs, more details can be found in [13] and the references therein and in [14]. Blind estimation of the ATFs is however beyond the scope of this paper in which we assume that the ATFs between the source and the sensors are known.

Manuscript received September 05, 2008; revised April 13, 2009. First published June 05, 2009; current version published October 23, 2009. The associate editor coordinating the review of this manuscript and approving it for publication was Dr. Jingdong Chen.

E. A. P. Habets is with the Department of Electrical and Electronic Engineering, Imperial College London, London SW7 2AZ, U.K. (e-mail: e.habets@imperial.ac.uk).

J. Benesty and J. Dmochowski are with the INRS-EMT, University of Quebec, Montreal, QC H5A 1K6, Canada.

I. Cohen is with the Technion, Israel Institute of Technology, Technion City, Haifa 32000, Israel.

S. Gannot is with the School of Engineering, Bar-Ilan University, Ramat Gan 52900, Israel.

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.2009.2024731

1558-7916/$26.00 © 2009 IEEE

In earlier works [12], it was observed that there is a tradeoff between the amount of speech dereverberation and noise reduction. However, this tradeoff was never rigorously analyzed.

Although the MVDR has attracted the attention of many researchers in the acoustics field [1], [2], [10]–[12], [15], [16] and has proven to be beneficial, a proper insight into its behavior in respect to its ability to reduce noise and to dereverberate the speech signal, is still lacking. A rigors analysis of this behavior is necessary to provide more insight. In addition, the results can be used to predict its performance in such an environment. In this paper, we study the MVDR beamformer in room acoustics. Specifically, the objectives of this paper are threefold: 1) we analyze the local and global behavior [1] of the MVDR beamformer, 2) we derive novel forms of the MVDR filter, and 3) we analyze the tradeoff between noise and reverberation reduction. The local and global behavior, as well as the tradeoff, is analyzed for different noise fields such as, for example, a mixture of coherent and non-coherent noise fields, entirely non-coherent noise fields and diffuse noise fields.

The paper is organized as follows. In Section II, the array model is formulated and the notation used in this paper is introduced. In Section III, we review the MVDR beamformer in the frequency domain and analyze the noise reduction performance. In Section IV, we define different performance measures that will be used in our analysis. In Section V, we analyze the performance of the MVDR beamformer. The performance evaluation that demonstrate the tradeoff between reverberation and noise reduction is presented in Section VI. Finally, conclusions are provided in Section VII.

II. Array Model

Consider the conventional signal model in which an N-element sensor array captures a convolved desired signal (speech source) in some noise field. The received signals are expressed as [1], [17]

$$\begin{aligned} y_n(k) &= g_n * s(k) + v_n(k) \\ &= x_n(k) + v_n(k), \quad n = 1, 2, \ldots, N \end{aligned} \tag{1}$$

where g_n is the impulse response from the unknown (desired) source $s(k)$ to the nth microphone, * stands for convolution, and $v_n(k)$ is the noise at microphone n. We assume that the signals $x_n(k)$ and $v_n(k)$ are uncorrelated and zero mean. All signals considered in this work are broadband. Without loss of generality, we consider the first microphone ($n = 1$) as the reference microphone. Our main objective in this paper is then to study the recovering of any one of the signals $x_1(k)$ (noise reduction only), $s(k)$ (total dereverberation and noise reduction), or a filtered version of $s(k)$ with the MVDR beamformer. Obviously, we can recover the reverberant component at one of the other microphones $x_2(k), \ldots, x_N(k)$. When we desire noise reduction, only the largest amount of noise reduction is attained by using the reference microphone with the highest signal to noise ratio.

In the frequency domain, (1) can be rewritten as

$$\begin{aligned} Y_n(j\omega) &= G_n(j\omega)S(j\omega) + V_n(j\omega) \\ &= X_n(j\omega) + V_n(j\omega), \quad n = 1, 2, \ldots, N \end{aligned} \tag{2}$$

where $Y_n(j\omega)$, $G_n(j\omega)$, $S(j\omega)$, $X_n(j\omega) = G_n(j\omega)S(j\omega)$, and $V_n(j\omega)$ are the discrete-time Fourier transforms (DTFTs) of $y_n(k)$, g_n, $s(k)$, $x_n(k)$, and $v_n(k)$, respectively, at angular frequency $\omega(-\pi < \omega \leq \pi)$ and j is the imaginary unit ($j^2 = -1$).

The N microphone signals in the frequency domain are better summarized in a vector notation as

$$\begin{aligned} \mathbf{y}(j\omega) &= \mathbf{g}(j\omega)S(j\omega) + \mathbf{v}(j\omega) \\ &= \mathbf{x}(j\omega) + \mathbf{v}(j\omega) \end{aligned} \tag{3}$$

where

$$\begin{aligned} \mathbf{y}(j\omega) &= [Y_1(j\omega) \quad Y_2(j\omega) \quad \cdots \quad Y_N(j\omega)]^T \\ \mathbf{x}(j\omega) &= [X_1(j\omega) \quad X_2(j\omega) \quad \cdots \quad X_N(j\omega)]^T \\ &= S(j\omega)[G_1(j\omega) \quad G_2(j\omega) \quad \cdots \quad G_N(j\omega)]^T \\ &= S(j\omega)\mathbf{g}(j\omega) \\ \mathbf{v}(j\omega) &= [V_1(j\omega) \quad V_2(j\omega) \quad \cdots \quad V_N(j\omega)]^T \end{aligned}$$

and superscript ^{T} denotes transpose of a vector or a matrix.

Using the power spectral density (PSD) of the received signal and the fact that $x_n(k)$ and $v_n(k)$ are uncorrelated, we get

$$\begin{aligned} \phi_{y_n}(j\omega) &= \phi_{x_n}(j\omega) + \phi_{v_n}(j\omega) \\ &= |G_n(j\omega)|^2\phi_s(j\omega) + \phi_{v_n}(j\omega), \\ &\qquad n = 1, 2, \ldots, N \end{aligned} \tag{4}$$

where $\phi_{y_n}(j\omega)$, $\phi_{x_n}(j\omega)$, $\phi_s(j\omega)$, and $\phi_{v_n}(j\omega)$ are the PSDs of the nth sensor input signal, the nth sensor reverberant speech signal, the desired signal, and the nth sensor noise signal, respectively.

The array processing, or beamforming, is then performed by applying a complex weight to each sensor and summing across the aperture

$$\begin{aligned} Z(j\omega) &= \mathbf{h}^H(j\omega)\mathbf{y}(j\omega) \\ &= \mathbf{h}^H(j\omega)[\mathbf{g}(j\omega)S(j\omega) + \mathbf{v}(j\omega)] \end{aligned} \tag{5}$$

where $Z(j\omega)$ is the beamformer output

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

is the beamforming weight vector which is suitable for performing spatial filtering at frequency ω, and superscript ^{H} denotes transpose conjugation of a vector or a matrix.

The PSD of the beamformer output is given by

$$\phi_z(j\omega) = \mathbf{h}^H(j\omega)\mathbf{\Phi}_x(j\omega)\mathbf{h}(j\omega) + \mathbf{h}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}(j\omega) \tag{6}$$

where

$$\begin{aligned} \mathbf{\Phi}_x(j\omega) &= E[\mathbf{x}(j\omega)\mathbf{x}^H(j\omega)] \\ &= \phi_s(j\omega)\mathbf{g}(j\omega)\mathbf{g}^H(j\omega) \end{aligned} \tag{7}$$

is the rank-one PSD matrix of the convolved speech signals with $E(\cdot)$ denoting mathematical expectation, and

$$\mathbf{\Phi}_v(j\omega) = E[\mathbf{v}(j\omega)\mathbf{v}^H(j\omega)] \tag{8}$$

is the PSD matrix of the noise field. In the rest of this paper, we assume that the noise is not fully coherent at the microphones so that $\mathbf{\Phi}_v(j\omega)$ is a full-rank matrix.

Now, we define a parameterized desired signal, which we denote by $Q(j\omega)S(j\omega)$, where $Q(j\omega)$ refers to a complex scaling factor that defines the nature of our desired signal. Let $G_1^{\mathrm{d}}(j\omega)$ denote the DTFT of the direct path response from the desired source to the first microphone. By setting $Q(j\omega) = G_1^{\mathrm{d}}(j\omega)$, we are stating that we desire both noise reduction and complete dereverberation. By setting $Q(j\omega) = G_1(j\omega)$, we are stating that we only desire noise reduction or in other words we desire to recover the reference sensor signal $X_1(j\omega) = G_1(j\omega)S(j\omega)$. In the following, we use the factor $Q(j\omega)$ in the definitions of performance measures and in the derivation of the MVDR beamformer.

III. Minimum Variance Distortionless Response Beamformer

We now derive the celebrated MVDR beamformer proposed by Capon [3] in the context of room acoustics.

Let us define the error signal between the output beamformer and the parameterized desired signal at frequency ω

$$\begin{aligned}\mathcal{E}(j\omega) &= Z(j\omega) - Q(j\omega)S(j\omega)\\ &= [\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) - Q(j\omega)]S(j\omega)\\ &\quad + \mathbf{h}^H(j\omega)\mathbf{v}(j\omega).\end{aligned} \tag{9}$$

The mean-squared error (MSE) is given by

$$\begin{aligned}J[\mathbf{h}(j\omega)] &= E[|\mathcal{E}(j\omega)|^2]\\ &= |\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) - Q(j\omega)|^2\phi_s(j\omega)\\ &\quad + \mathbf{h}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}(j\omega).\end{aligned} \tag{10}$$

This form of the MSE is helpful to derive the MVDR filter which is conceived by providing a fixed gain [in our case modeled by $Q(j\omega)$] to the signal while utilizing the remaining degrees of freedom to minimize the contribution of the noise and interference [second term of the right-hand side of (10)] to the array output[1]

$$\begin{aligned}\mathbf{h}_{\mathrm{MVDR}}(j\omega) = \arg\min_{\mathbf{h}(j\omega)} \mathbf{h}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}(j\omega)\\ \text{subject to } \mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = Q(j\omega).\end{aligned} \tag{11}$$

The solution to this constrained optimization problem is given by

$$\mathbf{h}_{\mathrm{MVDR}}(j\omega) = Q^*(j\omega)\frac{\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)}{\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)} \tag{12}$$

[1]The same MVDR filter can be found by minimizing $\mathbf{h}^H(j\omega)\mathbf{\Phi}_y(j\omega)\mathbf{h}(j\omega)$ subject to $\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = Q(j\omega)$ [18].

where superscript * denotes complex conjugation. In practice, the PSD matrix $\mathbf{\Phi}_v(j\omega)$ can be estimated during noise-only periods.

We can get rid of the explicit dependence on the above filter on the acoustic transfer functions $\{G_2(j\omega), \ldots, G_M(j\omega)\}$ by multiplying and dividing (12) by $\phi_s(j\omega)$ and using the fact that $\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega) = \mathrm{tr}[\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)\mathbf{g}^H(j\omega)]$ to obtain the following form

$$\mathbf{h}_{\mathrm{MVDR}}(j\omega) = \frac{Q^*(j\omega)}{G_1^*(j\omega)}\frac{\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)}{\mathrm{tr}\left[\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)\right]}\mathbf{u} \tag{13}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, and

$$\mathbf{u} = [1 \quad 0 \quad \cdots \quad 0 \quad 0]^T$$

is a vector of length N. Interestingly, we only need $G_1(j\omega)$ to achieve dereverberation and noise reduction.

Using the Woodbury's identity, another important form of the MVDR filter is derived

$$\mathbf{h}_{\mathrm{MVDR}}(j\omega) = C(j\omega)\mathbf{\Phi}_y^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)\mathbf{u} \tag{14}$$

where

$$C(j\omega) = \frac{Q^*(j\omega)}{G_1^*(j\omega)}\left\{1 + \frac{1}{\mathrm{tr}\left[\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)\right]}\right\} \tag{15}$$

and

$$\mathbf{\Phi}_y(j\omega) = E[\mathbf{y}(j\omega)\mathbf{y}^H(j\omega)] \tag{16}$$

is the PSD matrix of the microphone signals.

For the particular case, $Q(j\omega) = G_1(j\omega)$, where we only want to reduce the level of the noise (no dereverberation at all), we can get rid of the explicit dependence of the MVDR filter on all acoustic transfer functions to obtain the following forms [1]:

$$\begin{aligned}\mathbf{h}_{\mathrm{MVDR}}(j\omega) &= \frac{\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)}{\mathrm{tr}\left[\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)\right]}\mathbf{u}\\ &= \frac{\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_y(j\omega) - \mathbf{I}}{\mathrm{tr}\left[\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_y(j\omega)\right] - N}\mathbf{u}\end{aligned} \tag{17}$$

where $\mathbf{I}$ is the $N \times N$ identity matrix. Hence, noise reduction can be achieved without explicitly estimating the acoustic transfer functions.

IV. Performance Measures

In this section, we present some very useful measures that will help us better understand how noise reduction and speech dereverberation work with the MVDR beamformer in a real room acoustic environment.

To be consistent with prior works we define the local input signal-to-noise ratio (SNR) with respect to the parameterized desired signal [given by $Q(j\omega)S(j\omega)$] and the noise signal received by the first microphone, i.e.,

$$\mathrm{iSNR}[Q(j\omega)] = \frac{|Q(j\omega)|^2\phi_s(j\omega)}{\phi_{v_1}(j\omega)}, \quad \omega \in (-\pi, \pi] \tag{18}$$

where $\phi_{v_1}(j\omega)$ is the PSD of the noise signal $v_1(k)$. The global input SNR is given by

$$\text{iSNR}(Q) = \frac{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(j\omega) d\omega}{\int_{-\pi}^{\pi} \phi_{v_1}(j\omega) d\omega}. \tag{19}$$

After the MVDR beamforming operation with the frequency-domain model given in (6), the local output SNR is

$$\begin{aligned}\text{oSNR}[\mathbf{h}_{\text{MVDR}}(j\omega)] &= \frac{\left|\mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{g}(j\omega)\right|^2 \phi_s(j\omega)}{\mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}_{\text{MVDR}}(j\omega)} \\ &= \frac{|Q(j\omega)|^2 \phi_s(j\omega)}{\mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}_{\text{MVDR}}(j\omega)}.\end{aligned} \tag{20}$$

By substituting (12) in (20) it can be easily shown that

$$\begin{aligned}&\text{oSNR}[\mathbf{h}_{\text{MVDR}}(j\omega)] \\ &= \frac{|Q(j\omega)|^2 \phi_s(j\omega)}{|Q(j\omega)|^2 \frac{\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)}{\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)} \mathbf{\Phi}_v(j\omega) \frac{\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)}{\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)}} \\ &= \phi_s(j\omega)\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega) \\ &= \text{tr}[\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)], \quad \omega \in (-\pi, \pi].\end{aligned} \tag{21}$$

It is extremely important to observe that the desired scaling provided by $Q(j\omega)$ has no impact on the resulting local output SNR (but has an impact on the local input SNR). The global output SNR with the MVDR filter is

$$\begin{aligned}&\text{oSNR}(\mathbf{h}_{\text{MVDR}}) \\ &= \frac{\int_{-\pi}^{\pi} \left|\mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{g}(j\omega)\right|^2 \phi_s(j\omega) d\omega}{\int_{-\pi}^{\pi} \mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}_{\text{MVDR}}(j\omega) d\omega} \\ &= \frac{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(j\omega) d\omega}{\int_{-\pi}^{\pi} |Q(j\omega)|^2 [\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)]^{-1} d\omega} \\ &= \frac{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(j\omega) d\omega}{\int_{-\pi}^{\pi} \text{oSNR}^{-1}[\mathbf{h}_{\text{MVDR}}(j\omega)] |Q(j\omega)|^2 \phi_s(j\omega) d\omega}.\end{aligned} \tag{22}$$

Contrary to the local output SNR, the global output SNR depends strongly on the complex scaling factor $Q(j\omega)$.

Another important measure is the level of noise reduction achieved through beamforming. Therefore, we define the local noise-reduction factor as the ratio of the PSD of the original noise at the reference microphone over the PSD of the residual noise

$$\xi_{\text{nr}}[\mathbf{h}(j\omega)] = \frac{\phi_{v_1}(j\omega)}{\mathbf{h}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}(j\omega)} \tag{23}$$

$$= \frac{\text{oSNR}[\mathbf{h}(j\omega)]}{\text{iSNR}[Q(j\omega)]} \cdot \frac{|Q(j\omega)|^2}{|\mathbf{h}^H(j\omega)\mathbf{g}(j\omega)|^2}, \quad \omega \in (-\pi, \pi]. \tag{24}$$

We see that $\xi_{\text{nr}}[\mathbf{h}(j\omega)]$ is the product of two terms. The first one is the ratio of the output SNR over the input SNR at frequency ω while the second term represents the local distortion introduced by the beamformer $\mathbf{h}(j\omega)$. For the MVDR beamformer we have $|\mathbf{h}_{\text{MVDR}}^H(j\omega)\mathbf{g}(j\omega)|^2 = |Q(j\omega)|^2$. Therefore, we can further simplify (24)

$$\xi_{\text{nr}}[\mathbf{h}_{\text{MVDR}}(j\omega)] = \frac{\text{oSNR}[\mathbf{h}_{\text{MVDR}}(j\omega)]}{\text{iSNR}[Q(j\omega)]}, \quad \omega \in (-\pi, \pi]. \tag{25}$$

In this case, the local noise-reduction factor tells us exactly how much the output SNR is improved (or not) compared to the input SNR.

Integrating across the entire frequency range in the numerator and denominator of (23) yields the global noise-reduction factor

$$\begin{aligned}\xi_{\text{nr}}(\mathbf{h}) &= \frac{\int_{-\pi}^{\pi} \phi_{v_1}(j\omega) d\omega}{\int_{-\pi}^{\pi} \mathbf{h}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}(j\omega) d\omega} \\ &= \frac{\text{oSNR}(\mathbf{h})}{\text{iSNR}(Q)} \cdot \frac{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(j\omega) d\omega}{\int_{-\pi}^{\pi} |\mathbf{h}^H(j\omega)\mathbf{g}(j\omega)|^2 \phi_s(j\omega) d\omega}.\end{aligned} \tag{26}$$

The global noise-reduction factor is also the product of two terms. While the first one is the ratio of the global output SNR over the global input SNR, the second term is the global speech distortion due the beamformer. For the MVDR beamformer the global noise-reduction factor further simplifies to

$$\xi_{\text{nr}}(\mathbf{h}_{\text{MVDR}}) = \frac{\text{oSNR}(\mathbf{h}_{\text{MVDR}})}{\text{iSNR}(Q)}. \tag{27}$$

V. Performance Analysis

In this section, we analyze the performance of the MVDR beamformer and the tradeoff between the amount of speech dereverberation and noise reduction. When comparing the noise-reduction factor of different MVDR beamformers (with different objectives) it is of great importance that the comparison is conducted in a fair way. In Section V-A we will discuss this issue and propose a viable comparison method. In Sections V-B and V-C, we analyze the local and global behavior of the output SNR and the noise-reduction factor obtained by the MVDR beamformer, respectively. In addition, we analyze the tradeoff between dereverberation and noise reduction. In Sections V-D, V-E and V-F, we analyze the MVDR performance in three different noise fields, *viz.*, 1) non-coherent noise fields, 2) mixed coherent and non-coherent noise fields, and 3) diffuse noise fields.

Before we proceed we define the local squared Pearson correlation coefficient (SPCC) or magnitude squared coherence function (MSCF), which is the frequency-domain counterpart of the SPCC. In [19], the SPCC was used to analyze the noise reduction performance of the single-channel Wiener filter. Let $A(j\omega)$ and $B(j\omega)$ be the DTFTs of the two zero-mean real-valued random sequences a and b. Then the local SPCC between $A(j\omega)$ and $B(j\omega)$ at frequency ω is defined as

$$\begin{aligned}|\rho[A(j\omega), B(j\omega)]|^2 &= \frac{|E[A(j\omega)B^*(j\omega)]|^2}{E[|A(j\omega)|^2]E[|B(j\omega)|^2]} \\ &= \frac{|\phi_{ab}(j\omega)|^2}{\phi_a(j\omega)\phi_b(j\omega)}.\end{aligned} \tag{28}$$

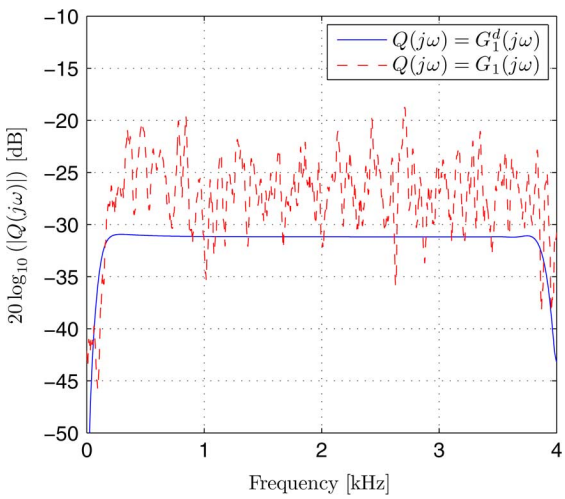

Fig. 1. Magnitude of the transfer functions $Q(j\omega) = \{G_1(j\omega), G_1^{\mathrm{d}}(j\omega)\}$ (reverberation time $T_{60} = 0.5$ s, source-receiver distance $D = 2.5$ m).

It is clear that the local SPCC always takes its values between 0 and 1.

A. On the Comparison of Different MVDR Beamformers

One of the main objectives of this work is to compare MVDR beamformers with different constraints. When we desire noise-reduction only, the constraint of the MVDR beamformer is given by $\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = G_1(j\omega)$. When we desire complete dereverberation and noise reduction we can use the constraint $\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = G_1^{\mathrm{d}}(j\omega)$, where $G_1^{\mathrm{d}}(j\omega)$ denotes the transfer function of the direct path response from the source to the first microphone. In Fig. 1, the magnitude of the transfer functions $G_1(j\omega)$ and $G_1^{\mathrm{d}}(j\omega)$ are shown. The transfer function $G_1(j\omega)$ was generated using the image-method [20], the distance between the source and the microphone was 2.5 m and the reverberation time was 500 ms. The transfer function $G_1^{\mathrm{d}}(j\omega)$ was obtained by considering only the direct path. As expected from a physical point of view, we can see that the energy of $G_1(j\omega)$ is larger than the energy of $G_1^{\mathrm{d}}(j\omega)$. In addition we observe that for some frequencies $|G_1(j\omega)|^2$ is smaller than $|G_1^{\mathrm{d}}(j\omega)|^2$. Evidently, the power of the desired signal $G_1^{\mathrm{d}}(j\omega)S(j\omega)$ is always smaller than the power of the desired signal $G_1(j\omega)S(j\omega)$.

Now let us first look at an illustrative example. Obviously, by choosing any constraint $\mathbf{h}_\gamma^H(j\omega)\mathbf{g}(j\omega) = \gamma \cdot G_1^{\mathrm{d}}(j\omega)(0 < \gamma \leq 1 \wedge \gamma \in \mathbb{R})$ we desire both noise reduction and complete dereverberation. Now the MVDR filters $\mathbf{h}_{\mathrm{MVDR},\gamma}(j\omega)$ are equal to $\gamma\mathbf{h}_{\mathrm{MVDR}}(j\omega)$, i.e., by scaling the desired signal we scale the MVDR filter. Consequently, we have also scaled the noise signal at the output. When we would directly calculate the noise-reduction factor of the beamformers using (25) we obtain different results since

$$\xi_{\mathrm{nr}}[\mathbf{h}_{\mathrm{MVDR},\gamma}(j\omega)] \neq \xi_{\mathrm{nr}}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)] \text{ for } \gamma \neq 1. \tag{29}$$

This can also be explained by the fact that the local output SNRs of all MVDR beamformers $\mathbf{h}_{\mathrm{MVDR},\gamma}(j\omega)$ are equal because the local output SNR [as defined in (20)] is independent of γ while the local input SNR [as defined in (18)] is dependent on γ. A similar problem occurs when we like to compare the noise-reduction factor of MVDR beamformers with completely different constraints because the power of the reverberant signal is much larger than the power of the direct sound signal. This abnormality can be corrected by normalizing the power of the output signal. Fundamentally, the definition of the MVDR beamformer depends on $Q(j\omega)$. Therefore, the choice of different desired signals [given by $Q(j\omega)S(j\omega)$] reflects the definition of the iSNR. Basically we can apply any normalization provided that the power of the desired signals at the output is equal. However, to obtain a meaningful noise-reduction factor and to be consistent with earlier works we propose to make the power of the desired signal at the output of the beamformer equal to the power of the signal that would be obtained when using the constraint $\mathbf{h}^H(j\omega)\mathbf{g}(j\omega) = G_1(j\omega)$. The global normalization factor $\eta(Q, G_1)$ is therefore given by

$$\eta(Q, G_1) = \sqrt{\frac{\int_{-\pi}^{\pi} |G_1(j\omega)|^2 \phi_s(j\omega) d\omega}{\int_{-\pi}^{\pi} |Q(j\omega)|^2 \phi_s(j\omega) d\omega}}. \tag{30}$$

B. Local Analyses

Let us first investigate the local behavior of the input and output SNRs via the SPCCs. Indeed, the local SPCC between the parameterized desired signal $Q(j\omega)S(j\omega)$ and the reference microphone signal $Y_1(j\omega)$ is

$$\begin{aligned} &|\rho[Q(j\omega)S(j\omega), Y_1(j\omega)]|^2 \\ &\quad = \frac{|G_1(j\omega)|^2 \phi_s(j\omega)}{\phi_{y_1}(j\omega)} \\ &\quad = \frac{\phi_s(j\omega)}{\phi_s(j\omega) + |G_1(j\omega)|^{-2}\phi_{v_1}(j\omega)}. \end{aligned} \tag{31}$$

Expression (31) tells us how much the signals $Q(j\omega)S(j\omega)$ and $Y_1(j\omega)$ are coherent at frequency ω, i.e., how noisy the reference microphone signal is. In addition, we note that the local SPCC (31) does not depend on the complex scaling factor $Q(j\omega)$.

At the same time, the local SPCC between the parameterized desired signal, $Q(j\omega)S(j\omega)$, and the beamformer output $\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)$ is maximized by $\mathbf{h}_{\mathrm{MVDR}}(j\omega)$ and does not depend on $Q(j\omega)$ [the same way the local output SNR does not depend on $Q(j\omega)$]

$$\begin{aligned} &\left|\rho\left[Q(j\omega)S(j\omega), \mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2 \\ &\quad = \frac{|Q(j\omega)|^2 \phi_s(j\omega)}{\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{\Phi}_y(j\omega)\mathbf{h}_{\mathrm{MVDR}}(j\omega)} \\ &\quad = \frac{|Q(j\omega)|^2 \phi_s(j\omega)}{|Q(j\omega)|^2 \phi_s(j\omega) + \mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{\Phi}_v(j\omega)\mathbf{h}_{\mathrm{MVDR}}(j\omega)} \\ &\quad = \frac{\phi_s(j\omega)}{\phi_s(j\omega) + \left(\mathbf{g}^H(j\omega)\mathbf{\Phi}_v^{-1}(j\omega)\mathbf{g}(j\omega)\right)^{-1}}. \end{aligned} \tag{32}$$

Indeed, (32) is equal to one when $\mathbf{\Phi}_v(j\omega)$ approaches zero and is equal to zero when $\phi_s(j\omega)$ equals zero.

The most important goal of a beamforming algorithm is to improve the local SNR after filtering. Therefore, we must design the beamforming weight vectors, $\mathbf{h}(j\omega), \omega \in (-\pi, \pi]$, in such a way that $\mathrm{oSNR}[\mathbf{h}(j\omega)] \geq \mathrm{iSNR}[Q(j\omega)]$. We next give an

interesting property that will give more insights into the local SNR behavior of the MVDR beamformer.

Property 5.1: With the MVDR filter given in (12), the local output SNR times $|Q(j\omega)|^2$ is always greater than or equal to the local input SNR times $|G_1(j\omega)|^2$, i.e.,

$$|Q(j\omega)|^2 \cdot \mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)] \geq |G_1(j\omega)|^2 \cdot \mathrm{iSNR}[Q(j\omega)], \quad \forall\omega \tag{33}$$

which can also be expressed using (18) as

$$\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)] \geq \frac{|G_1(j\omega)|^2\phi_s(j\omega)}{\phi_{v_1}(j\omega)}, \quad \forall\omega. \tag{34}$$

Proof: See Appendix A. ■

The normalized local noise-reduction factor is defined as

$$\begin{aligned}\tilde{\xi}_{\mathrm{nr}}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)] &= \frac{1}{\eta^2(Q,G_1)}\frac{\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]}{\mathrm{iSNR}[Q(j\omega)]} \\ &= \frac{1}{\eta^2(Q,G_1)|Q(j\omega)|^2}\cdot\frac{\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]\cdot\phi_{v_1}(j\omega)}{\phi_s(j\omega)} \\ &= \frac{1}{\zeta[Q(j\omega),G_1(j\omega)]}\cdot\frac{\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]\cdot\phi_{v_1}(j\omega)}{\phi_s(j\omega)}\end{aligned} \tag{35}$$

where $\zeta[Q(j\omega),G_1(j\omega)] = \eta^2(Q,G_1)|Q(j\omega)|^2$. Indeed, for different MVDR beamformers the noise-reduction factor varies due to $\zeta[Q(j\omega),G_1(j\omega)]$, since the local output SNR $\phi_{v_1}(j\omega)$ and $\phi_s(j\omega)$ do not depend on $Q(j\omega)$. Since $\zeta[Q(j\omega),G_1(j\omega)] = \zeta[\gamma Q(j\omega),G_1(j\omega)](0 < \gamma \leq 1)$ the normalized local noise-reduction factor is independent of the global scaling factor γ.

To gain more insight into the local behavior of $\zeta[Q(j\omega),G_1(j\omega)]$ we analyzed several acoustic transfer functions. To simplify the following discussion we assume that the power spectral density $\phi_s(j\omega) = 1$ for all ω. Let us decompose the transfer function $G_1(j\omega)$ into two parts. The first part $G_1^{\mathrm{d}}(j\omega)$ is the DTFT the direct path, while the second part $G_1^{\mathrm{r}}(j\omega)$ is the DTFT of the reverberant part. Now let us define the desired response as

$$Q(j\omega,\alpha) = G_1^{\mathrm{d}}(j\omega) + \alpha\, G_1^{\mathrm{r}}(j\omega) \tag{36}$$

where the parameter $0 \leq \alpha \leq 1$ controls the direct-to-reverberation ratio (DRR) of the desired response. In Fig. 2(a), we plotted $\zeta[Q(j\omega,\alpha),G_1(j\omega)]$ for $\alpha = \{0, 0.2, 1\}$. Due to the normalization the energy of $\zeta[Q(j\omega,\alpha),G_1(j\omega)]$ (and therefore its mean value) does not depend on α. Locally, we can see that the deviation with respect to $|G_1^{\mathrm{d}}(j\omega)|^2$ increases when α increases (i.e., when the DRR decreases). In Fig. 2(b), we plotted the histogram of $\zeta[Q(j\omega,\alpha),G_1(j\omega)]$ for $\alpha = \{0, 0.2, 1\}$. First, we observe that the probability that $\zeta[Q(j\omega,\alpha),G_1(j\omega)]$ is smaller than its mean value decreases when α decreases (i.e., when the DRR increases). Second, we observe that the distribution is stretched out towards negative values on the decibel's logarithmic scale when α increases.

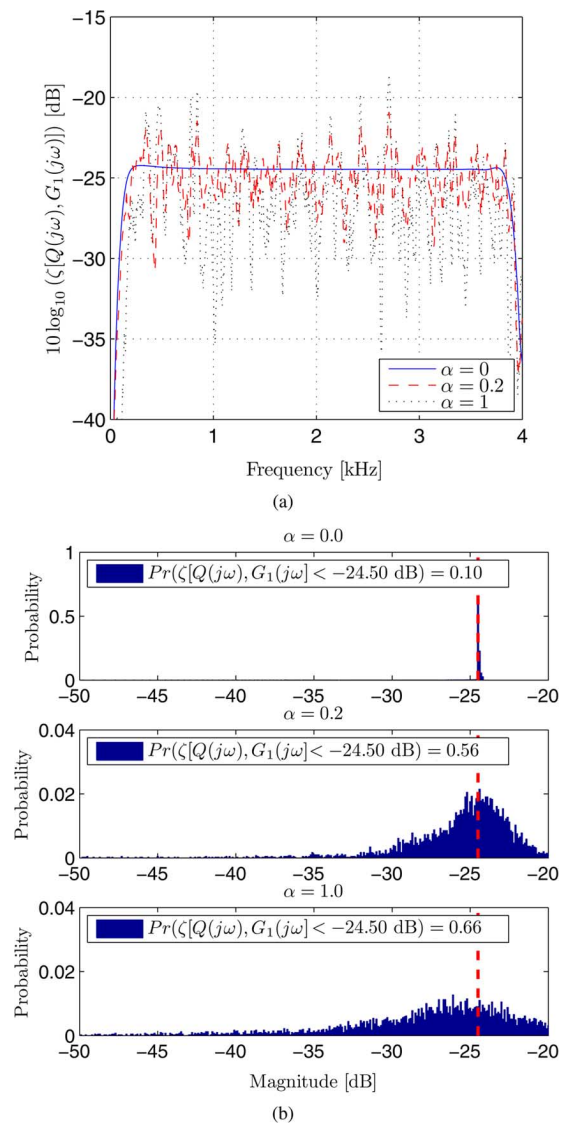

Fig. 2. (a) Normalized transfer functions $\zeta[Q(j\omega,\alpha),G_1(j\omega)]$ with $Q(j\omega,\alpha) = G_1^{\mathrm{d}}(j\omega) + \alpha\, G_1^{\mathrm{r}}(j\omega)$ for $\alpha = \{0, 0.2, 1\}$. (b) Histograms of $10\log_{10}(\zeta[Q(j\omega,\alpha),G_1(j\omega)])$.

Hence, when the desired speech signal contains less reverberation it is more likely that $\zeta[Q(j\omega,\alpha),G_1(j\omega)]$ will increase and that the local noise-reduction factor will decrease. Therefore, it is likely that the highest local noise reduction is achieved when we desire only noise reduction, i.e., for $Q(j\omega) = G_1(j\omega)$.

Using Property 5.1, we deduce a lower bound for the normalized local noise-reduction factor

$$\tilde{\xi}_{\mathrm{nr}}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)] \geq \frac{1}{\eta^2(Q,G_1)|Q(j\omega)|^2}\,|G_1(j\omega)|^2. \tag{37}$$

For $Q(j\omega) = G_1(j\omega)$ we obtain

$$\tilde{\xi}_{\mathrm{nr}}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)] \geq 1. \tag{38}$$

Expression (38) proves that there is always noise-reduction when we desire only noise reduction. However, in other situations we cannot guarantee that there is noise reduction.

C. Global Analyses

Using (27), (22), and (19), we deduce the normalized global noise-reduction factor

$$\begin{aligned}\tilde{\xi}_{\rm nr}(\mathbf{h}_{\rm MVDR}) &= \frac{1}{\eta^2(Q,G_1)}\frac{\rm oSNR(\mathbf{h}_{MVDR})}{\rm iSNR(Q)}\\ &= \frac{1}{\eta^2(Q,G_1)}\frac{\int_{-\pi}^{\pi}\phi_{v_1}(j\omega)d\omega}{\int_{-\pi}^{\pi}{\rm oSNR}^{-1}[\mathbf{h}_{\rm MVDR}(j\omega)]|Q(j\omega)|^2\phi_s(j\omega)d\omega}\\ &= \frac{\int_{-\pi}^{\pi}\phi_{v_1}(j\omega)d\omega}{\int_{-\pi}^{\pi}{\rm oSNR}^{-1}[\mathbf{h}_{\rm MVDR}(j\omega)]\zeta[Q(j\omega),G_1(j\omega)]\phi_s(j\omega)d\omega}.\end{aligned} \tag{39}$$

This normalized global noise-reduction factor behaves, with respect to $Q(j\omega)$, similarly to its local counterpart. It can easily be verified that the normalized global noise-reduction factor for $\gamma\cdot Q(j\omega)$ is independent of γ. Due to the complexity of (39) it is difficult to predict the exact behavior of the normalized global noise-reduction factor. From the analyses in the previous subsection we do know that the distribution of $\zeta[Q(j\omega),G_1(j\omega)]$ is stretched out towards zero when the DRR decreases. Hence, for each frequency it is likely that $\zeta[Q(j\omega),G_1(j\omega)]$ will decrease when the DRR decreases. Consequently, we expect that the normalized global noise-reduction factor will always increase when the DRR decreases. The expected behavior of the normalized global noise-reduction factor is consistent with the results presented in Section VI.

D. Non-Coherent Noise Field

Let us assume that the noise field is homogeneous and spatially white. In case the noise variance at each microphone is equal to $\sigma_{\rm nc}^2(j\omega)$ the noise covariance matrix $\mathbf{\Phi}_v(j\omega)$ simplifies to $\sigma_{\rm nc}^2(j\omega)\mathbf{I}$. In the latter case, the MVDR beamformer simplifies to

$$\mathbf{h}_{\rm MVDR}(j\omega) = Q^*(j\omega)\frac{\mathbf{g}(j\omega)}{\|\mathbf{g}(j\omega)\|^2} \tag{40}$$

where $\|\mathbf{g}(j\omega)\|^2 = \mathbf{g}^H(j\omega)\mathbf{g}(j\omega)$. For $Q(j\omega)=G_1(j\omega)$ this is the well-known matched beamformer [21], which generalizes the delay-and-sum beamformer. The local output SNR and normalized local noise-reduction factor can be deduced by substituting $\sigma_{\rm nc}^2(j\omega)\mathbf{I}$ in (21) and (35), and results in

$$\begin{aligned}{\rm oSNR}[\mathbf{h}_{\rm MVDR}(j\omega)] &= {\rm tr}\left[\sigma_{\rm nc}^{-2}(j\omega)\mathbf{\Phi}_x(j\omega)\right]\\ &= \sigma_{\rm nc}^{-2}(j\omega)\|\mathbf{g}(j\omega)\|^2\phi_s(j\omega).\end{aligned} \tag{41}$$

and

$$\tilde{\xi}_{\rm nr}[\mathbf{h}_{\rm MVDR}(j\omega)] = \frac{1}{\eta^2(Q,G_1)|Q(j\omega)|^2}\|\mathbf{g}(j\omega)\|^2 \tag{42}$$

respectively. When $Q(j\omega)=G_1(j\omega)$ the normalization factor $\eta(Q,G_1)$ equals 1, the normalized noise-reduction factor then becomes

$$\begin{aligned}\tilde{\xi}_{\rm nr}[\mathbf{h}_{\rm MVDR}(j\omega)] &= \frac{\|\mathbf{g}(j\omega)\|^2}{|G_1(j\omega)|^2}\\ &= \left(1+\sum_{n=2}^{N}\frac{|G_n(j\omega)|^2}{|G_1(j\omega)|^2}\right).\end{aligned} \tag{43}$$

As we expected from (38), the normalized noise-reduction factor is always larger than 1 when $Q(j\omega)=G_1(j\omega)$. However, in other situations we cannot guarantee that there is noise reduction.

The normalized global noise-reduction factor is given by

$$\begin{aligned}\tilde{\xi}_{\rm nr}(\mathbf{h}_{\rm MVDR}) &= \frac{1}{\eta^2(Q,G_1)}\frac{\int_{-\pi}^{\pi}\sigma_{\rm nc}^{-2}(j\omega)\,\|\mathbf{g}(j\omega)\|^2\,\phi_s(j\omega)\,d\omega}{\int_{-\pi}^{\pi}\sigma_{\rm nc}^{-2}(j\omega)|Q(j\omega)|^2\phi_s(j\omega)\,d\omega}\\ &= \frac{\int_{-\pi}^{\pi}\sigma_{\rm nc}^{-2}(j\omega)\,\|\mathbf{g}(j\omega)\|^2\phi_s(j\omega)\,d\omega}{\int_{-\pi}^{\pi}\sigma_{\rm nc}^{-2}(j\omega)\zeta[Q(j\omega),G_1(j\omega)]\phi_s(j\omega)d\omega}.\end{aligned} \tag{44}$$

In an anechoic environment where the source is positioned in the far-field of the array, $G_n(j\omega)$ are steering vectors and $|Q(j\omega)|^2 = |G_n(j\omega)|^2, \quad \forall n$. In this case the normalized global noise-reduction factor simplifies to

$$\tilde{\xi}_{\rm nr}(\mathbf{h}_{\rm MVDR}) = N. \tag{45}$$

The latter results in consistent with earlier works and shows that the noise-reduction factor only depends on the number of microphones. When the PSD matrices of the noise and microphone signals are known we can compute the MVDR filter using (17), i.e., we do not require any *a prior* knowledge of the direction of arrival.

E. Coherent Plus Non-Coherent Noise Field

Let $\mathbf{d}(j\omega) = [D_1(j\omega)\,D_2(j\omega)\cdots D_N(j\omega)]^T$ denote the ATFs between a noise source and the array. The noise covariance matrix can be written as

$$\mathbf{\Phi}_v(j\omega) = \sigma_{\rm c}^2(j\omega)\mathbf{d}(j\omega)\mathbf{d}^H(j\omega) + \sigma_{\rm nc}^2(j\omega)\mathbf{I}. \tag{46}$$

Using Woodbury's identity we have

$$\begin{aligned}\mathbf{\Phi}_v^{-1}(j\omega) &= \left(\sigma_{\rm c}^2(j\omega)\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)+\sigma_{\rm nc}^2(j\omega)\mathbf{I}\right)^{-1}\\ &= \frac{1}{\sigma_{\rm nc}^2(j\omega)}\mathbf{I} - \frac{1}{\sigma_{\rm nc}^2(j\omega)}\mathbf{d}(j\omega)\\ &\quad\times\left(\frac{1}{\sigma_{\rm c}^2(j\omega)}+\mathbf{d}^H(j\omega)\frac{1}{\sigma_{\rm nc}^2(j\omega)}\mathbf{d}(j\omega)\right)^{-1}\\ &\quad\cdot\mathbf{d}^H(j\omega)\frac{1}{\sigma_{\rm nc}^2(j\omega)}\\ &= \frac{1}{\sigma_{\rm nc}^2(j\omega)}\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\frac{\sigma_{\rm nc}^2(j\omega)}{\sigma_{\rm c}^2(j\omega)}+\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right).\end{aligned} \tag{47}$$

Now, the MVDR beamformer becomes

$$\mathbf{h}_{\mathrm{MVDR}}(j\omega)=Q^*(j\omega)\frac{\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\frac{\sigma^2_{\mathrm{nc}}(j\omega)}{\sigma^2_{\mathrm{c}}(j\omega)}+\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right)\mathbf{g}(j\omega)}{\mathbf{g}^H(j\omega)\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\frac{\sigma^2_{\mathrm{nc}}(j\omega)}{\sigma^2_{\mathrm{c}}(j\omega)}+\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right)\mathbf{g}(j\omega)}. \tag{48}$$

The local output SNR and normalized local noise-reduction factor are given by

$$\begin{aligned}&\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]\\&\quad=\frac{1}{\sigma^2_{\mathrm{nc}}(j\omega)}\\&\qquad\cdot\mathrm{tr}\left[\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\frac{\sigma^2_{\mathrm{nc}}(j\omega)}{\sigma^2_{\mathrm{c}}(j\omega)}+\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right)\mathbf{\Phi}_x(j\omega)\right].\end{aligned} \tag{49}$$

and

$$\begin{aligned}&\tilde{\xi}_{\mathrm{nr}}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]\\&\quad=\frac{1}{\eta^2(Q,G_1)|Q(j\omega)|^2}\\&\qquad\cdot\frac{\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]\cdot\phi_{v_1}(j\omega)}{\phi_s(j\omega)}\\&\quad=\frac{1}{\zeta[Q(j\omega),G_1(j\omega)]\,\phi_s(j\omega)}\\&\qquad\cdot\mathrm{tr}\left[\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\frac{\sigma^2_{\mathrm{nc}}(j\omega)}{\sigma^2_{\mathrm{c}}(j\omega)}+\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right)\mathbf{\Phi}_x(j\omega)\right]\\&\qquad\cdot\left(1+\frac{\sigma^2_{\mathrm{c}}(j\omega)}{\sigma^2_{\mathrm{nc}}(j\omega)}|D_1(j\omega)|^2\right).\end{aligned} \tag{50}$$

The noise reduction depends on the ratio between the variance of the non-coherent and coherent, and on the inner product of $\mathbf{d}(j\omega)$ and $\mathbf{g}(j\omega)$ [22].

Obviously, the noise covariance matrix $\mathbf{\Phi}_v(j\omega)$ needs to be full-rank. However, from a theoretical point of view we can analyze the coherent noise at the output of the MVDR beamformer [given by $\mathbf{h}^H_{\mathrm{MVDR}}(j\omega)\mathbf{d}(j\omega)\sigma_{\mathrm{c}}(j\omega)$] when the ratio $(\sigma^2_{\mathrm{nc}}(j\omega))/(\sigma^2_{\mathrm{c}}(j\omega))$ approaches zero, i.e., the noise field becomes more and more coherent. Provided that $\mathbf{d}(j\omega)\neq\mathbf{g}(j\omega)$ the coherent noise at the output of the beamformer is given by

$$\begin{aligned}&\lim_{\frac{\sigma^2_{\mathrm{nc}}(j\omega)}{\sigma^2_{\mathrm{c}}(j\omega)}\to 0}\mathbf{h}^H_{\mathrm{MVDR}}(j\omega)\mathbf{d}(j\omega)\sigma_{\mathrm{c}}(j\omega)\\&\quad=Q^*(j\omega)\frac{\mathbf{g}^H(j\omega)\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right)\mathbf{d}(j\omega)}{\mathbf{g}^H(j\omega)\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right)\mathbf{g}(j\omega)}\sigma_{\mathrm{c}}(j\omega)\\&\quad=Q^*(j\omega)\frac{\mathbf{g}^H\left(\mathbf{d}(j\omega)-\mathbf{d}(j\omega)\right)}{\mathbf{g}^H\left(\mathbf{I}-\frac{\mathbf{d}(j\omega)\mathbf{d}^H(j\omega)}{\mathbf{d}^H(j\omega)\mathbf{d}(j\omega)}\right)\mathbf{g}(j\omega)}\sigma_{\mathrm{c}}(j\omega)\\&\quad=0.\end{aligned}$$

For $\mathbf{d}(j\omega)=\mathbf{g}(j\omega)$ there is a contradiction, since the desired signal and the noise come from the same point.

F. Diffuse Noise Field

In highly reverberant acoustical environment, such as a car enclosure, the noise field tends to be diffused (see for instance [23], [24]). A diffused noise field consists of infinite independent noise sources that are equi-distributed on a sphere around the array. The local PCC between signals received by two sensors (i,j) with distance d_{ij} can be found in [23], and is given in the following expression:

$$\begin{aligned}\rho[V_i(j\omega),V_j(j\omega)]&=\frac{\phi_{v_iv_j}(j\omega)}{\sqrt{\phi_{v_i}(j\omega)\phi_{v_j}(j\omega)}}\\&=\frac{\sin(\omega d_{ij}/c)}{\omega d_{ij}/c}\end{aligned} \tag{51}$$

where c denotes the sound velocity. As can be seen from (51), the coherence between the sensors decreases rapidly when the frequency increases. The coherence matrix is given by

$$\begin{aligned}&\mathbf{\Gamma}_v(j\omega)\\&\quad=\begin{bmatrix}1&\rho[V_1,V_2(j\omega)]&\cdots&\rho[V_1,V_N]\\\rho[V_2,V_1]&1&\cdots&\rho[V_2,V_N]\\\vdots&\vdots&\ddots&\vdots\\\rho[V_N,V_1]&\rho[V_N,V_2]&\cdots&1\end{bmatrix}.\end{aligned} \tag{52}$$

If $\sigma^2_{\mathrm{dn}}(j\omega)$ denotes the variance of the diffuse noise, then the noise covariance matrix is given by

$$\mathbf{\Phi}_v(j\omega)=\sigma^2_{\mathrm{dn}}(j\omega)\,\mathbf{\Gamma}_v(j\omega). \tag{53}$$

The local output SNR is given by[2]

$$\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]=\sigma^{-2}_{\mathrm{dn}}(j\omega)\mathrm{tr}\left[\mathbf{\Gamma}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)\right]. \tag{54}$$

The normalized local noise-reduction factor is given by

$$\begin{aligned}&\tilde{\xi}_{\mathrm{nr}}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]\\&\quad=\frac{1}{\eta^2(Q,G_1)}\frac{\mathrm{tr}\left[\mathbf{\Gamma}_v^{-1}(j\omega)\mathbf{\Phi}_x(j\omega)\right]}{|Q(j\omega)|^2\phi_s(j\omega)}\\&\quad=\frac{1}{\zeta[Q(j\omega),G_1(j\omega)]}\mathrm{tr}\left[\mathbf{\Gamma}_v^{-1}(j\omega)\mathbf{g}(j\omega)\mathbf{g}^H(j\omega)\right]\\&\quad=\frac{1}{\zeta[Q(j\omega),G_1(j\omega)]}\,\mathbf{g}^H(j\omega)\mathbf{\Gamma}_v^{-1}(j\omega)\mathbf{g}(j\omega).\end{aligned} \tag{55}$$

VI. Performance Evaluation

In this section, we evaluate the performance of the MVDR beamformer in room acoustics. We will demonstrate the tradeoff between speech dereverberation and noise reduction by computing the normalized noise-reduction factor in various

[2]For $\omega=0$ the diffuse noise field is entirely coherent, i.e., the rank of $\mathbf{\Phi}_v(j\omega)$ equals one. Consequently, the MVDR filter does not exist. However, in practice there is always an additional non-coherent noise term which makes the covariance matrix full-rank.

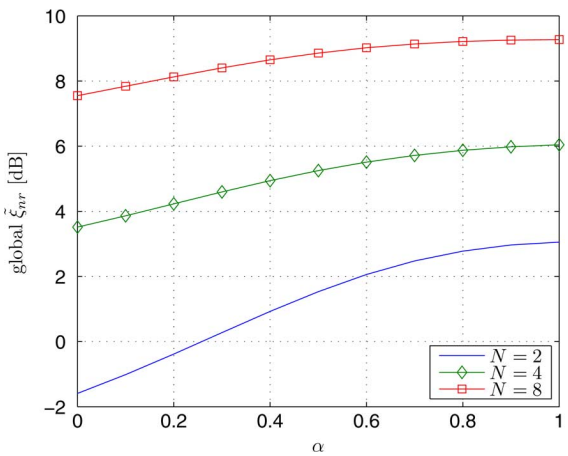

Fig. 3. Normalized global noise-reduction factor obtained using $N = \{2, 4, 8\}$ ($T_{60} = 0.3$, $D = 2$ m, non-coherent noise iSNR $= 5$ dB).

scenarios. A linear microphone array was used with two to eight microphones and an inter-microphone distance of 5 cm. The room size is $5 \times 4 \times 6$ m (length $\times$ width $\times$ height), the reverberation time of the enclosure varies between 0.2 to 0.4 s. All room impulse responses are generated using the image-method proposed by Allen and Berkley [20] with some necessary modifications that ensure proper inter-microphone phase delays as proposed by Peterson [25]. The distance between the desired source and the first microphone varies from 1 to 3 m. The desired source consists of speech like noise (USASI). The noise consists of a simple AR(1) process (autoregressive process of order one) that was created by filtering a stationary zero-mean Gaussian sequences with a linear time-invariant filter. We used non-coherent noise, a mixture of non-coherent noise and a coherent noise source, and diffuse noise.

In order to study the tradeoff more carefully, we need to control the amount of reverberation reduction. Here we propose to control the amount of reverberation reduction by changing the DRR of the desired response $Q(j\omega)$. As proposed in Section V-A, we control the DRR using the parameter $\alpha (0 \leq \alpha \leq 1)$. The complex scaling factor $Q(j\omega, \alpha)$ is calculated using (36). When the desired response equals $Q(j\omega, 0) = G_1^{\mathrm{d}}(j\omega)$, we desire both noise reduction and complete dereverberation. However, when the desired response equals $Q(j\omega, 1) = G_1(j\omega)$ we desire only noise reduction.

A. Influence of the Number of Microphones

In this section, we study the influence of the number of microphones used. The reverberation time was set to $T_{60} = 0.3$ s and the distance between the source and the first microphone was $D = 2$ m. The noise field is non-coherent and the global input SNR [for $Q(j\omega, 0) = G_1^{\mathrm{d}}(j\omega)$] was iSNR $= 5$ dB. In this experiment two, four, or eight microphones were used. In Fig. 3 the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$. First, we observe that there is a tradeoff between speech dereverberation and noise reduction. The largest amount of noise reduction is achieved for $\alpha = 1$, i.e., when no dereverberation is performed. While a smaller amount of noise reduction is achieved for $\alpha = 0$, i.e., when complete dereverberation is performed. In the case of two microphones ($N = 2$), we amplify the noise when we desire to complete dereverberate the speech signal. Second, we observe that the amount of noise reduction increases with approximately 3 dB if we double the number of microphones. Finally, we observe that the tradeoff becomes less evident when more microphones are used. When more microphones are available the degrees of freedom of the MVDR beamformer increases. In such a case the MVDR beamformer is apparently able to perform speech dereverberation without significantly sacrificing the amount of noise reduction.

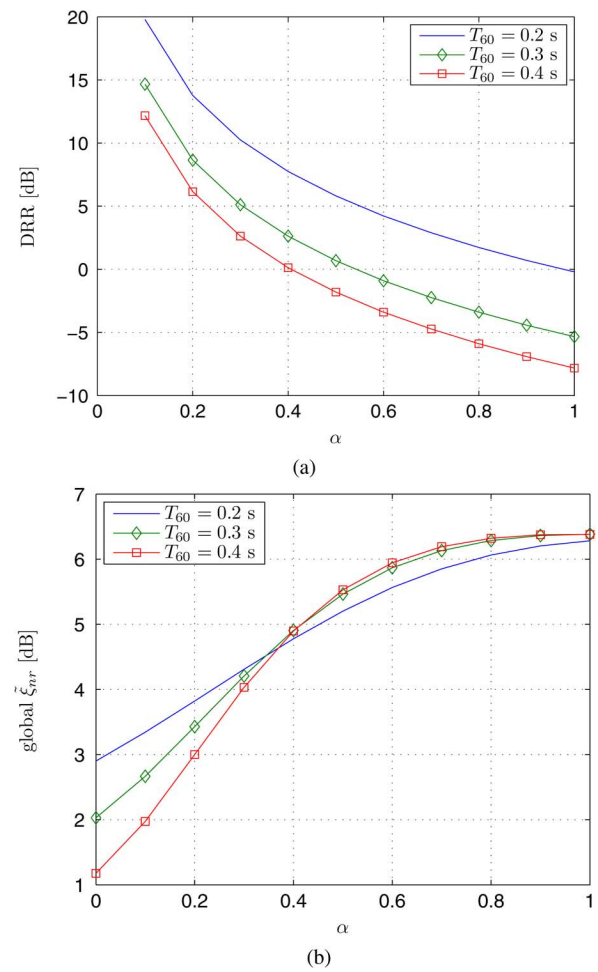

Fig. 4. (a) DRR of $Q(j\omega, \alpha)$ for $T_{60} = \{0.2, 0.3, 0.4\}$ s. (b) The normalized global noise-reduction factor obtained using $T_{60} = \{0.2, 0.3, 0.4\}$ s ($N = 4$, $D = 4$ m, non-coherent noise iSNR $= 5$ dB).

B. Influence of the Reverberation Time

In this section, we study the influence of the reverberation time. The distance between the source and the first microphone was set to $D = 4$ m. The noise field is non-coherent and the global input SNR [for $Q(j\omega) = G_1^{\mathrm{d}}(j\omega)$] was iSNR $= 5$ dB. In this experiment, four microphones were used, and the reverberation time was set to $T_{60} = \{0.2, 0.3, 0.4\}$ s. The DRR ratio of the desired response $Q(j\omega)$ is shown in Fig. 4(a). In Fig. 4(b), the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$. Again, we observe that there is a tradeoff between speech dereverberation and noise reduction. This experiment also shows that almost no noise reduction is sacrificed when we desire to increase the DRR to approximately -5 dB

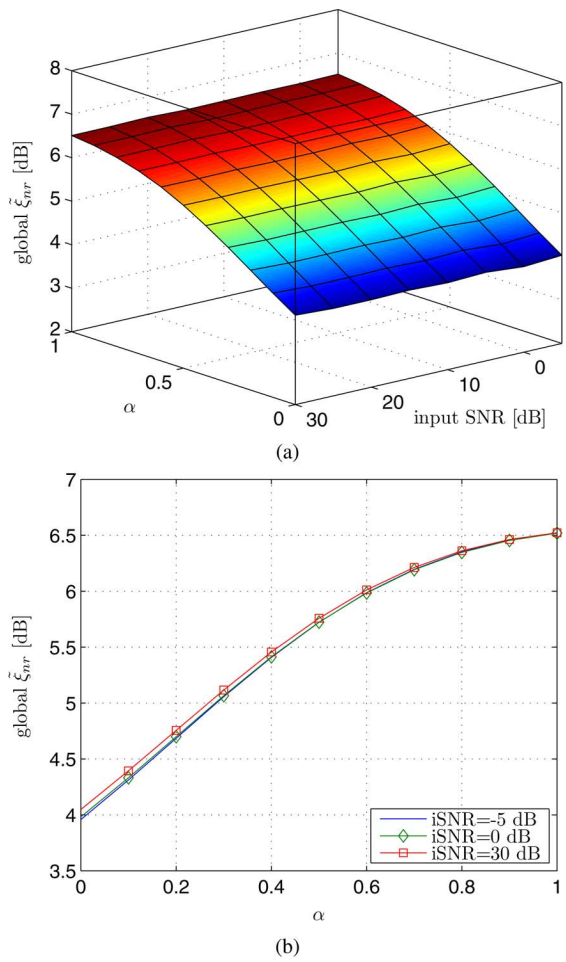

Fig. 5. Normalized global noise-reduction factor obtained using non-coherent noise iSNR $= \{-5, \ldots, 30\}$ dB ($T_{60} = 0.3$ s, $N = 4$, $D = 2$ m).

for $T_{60} \leq 0.3$ s. In other words, as long as the reverberant part of the signal is dominant (DRR ≤ -5 dB) we can reduce reverberation and noise without sacrificing too much noise reduction. However, when the DRR is increased further (DRR > -5 dB) the noise-reduction decreases.

C. Influence of the Noise Field

In this section we evaluate the normalized noise-reduction factor in various noise fields and study the tradeoff between noise reduction and dereverberation.

1) Non-Coherent Noise Field: In this section, we study the amount of noise reduction in a non-coherent noise field with different input SNRs. The distance between the source and the first microphone was set to $D = 2$ m. In this experiment four microphones were used, and the reverberation time was set to $T_{60} = 0.3$ s. In Fig. 5(a), the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$ and different input SNRs ranging from -5 dB to 30 dB. In Fig. 5(b), the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$ and input SNRs of -5, 0, and 30 dB. We observe the tradeoff between speech dereverberation and noise reduction as before. As expected from (44), for a non-coherent noise field the normalized global noise-reduction factor is independent of the input SNR. In Fig. 6, we depicted the normalized global noise-reduction factor for $\alpha = 0$ (i.e., complete dereverberation and noise reduction) and $\alpha = 1$ (i.e., noise reduction only) for different distances. It should be noted that the DRR is not monotonically decreasing with the distance. Therefore, the noise-reduction factor is not monotonically decreasing with the distance. Here four microphones were used and the reverberation time equals 0.3 s.

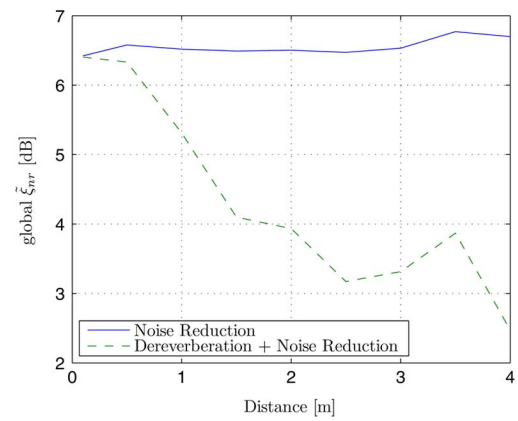

Fig. 6. Normalized global noise-reduction factor for one specific source trajectory obtained using $D = \{0.1, 0.5, 1, \ldots, 4\}$ m ($T_{60} = 0.3$ s, $N = 4$, non-coherent noise iSNR $= 5$ dB).

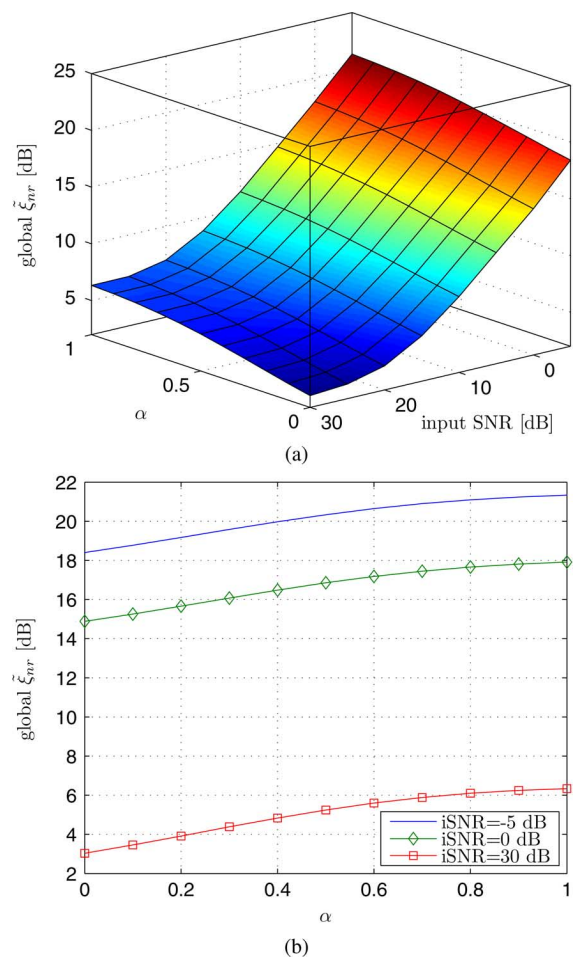

Fig. 7. Normalized global noise-reduction factor obtained using a coherent plus non-coherent noise iSNR$_{\mathrm{c}} = \{-5, \ldots, 30\}$ dB (iSNR$_{\mathrm{nc}} = 20$ dB, $T_{60} = 0.3$ s, $N = 4$, $D = 2$ m).

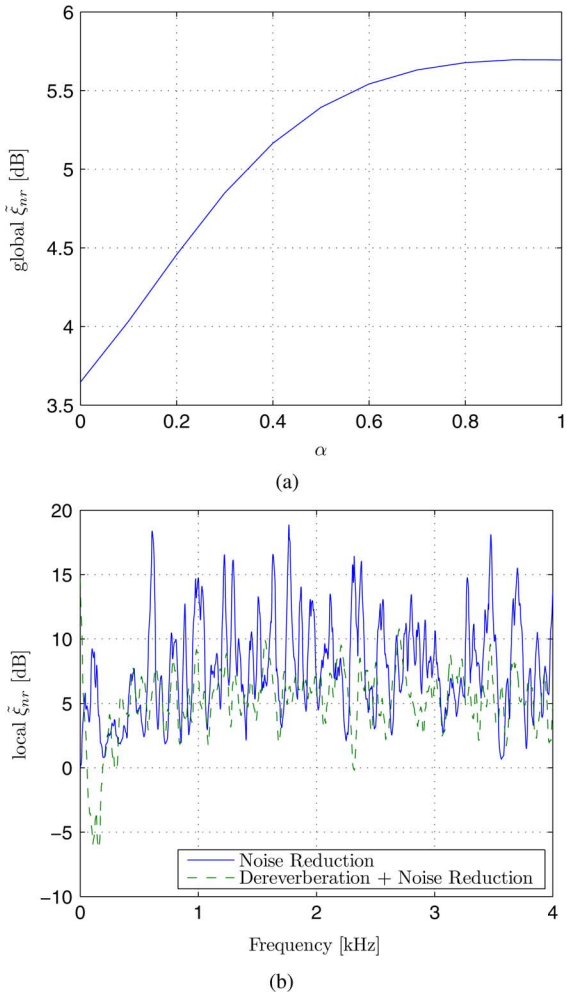

Fig. 8. Normalized noise-reduction factor obtained using a diffuse plus non-coherent noise field. (a) Global noise-reduction factor for $0 \leq \alpha \leq 1$, and (b) local noise-reduction factor for $\alpha = \{0, 1\}$ (iSNR$_{dn} = 0$ dB, iSNR$_{\rm nc} = 30$ dB, $T_{60} = 0.3$ s, $N = 4$, $D = 4$ m).

When we desire only noise reduction, the noise reduction is independent of the distance between the source and the first microphone. However, when we desire both dereverberation and noise reduction we see that the normalized global noise-reduction factor decreases rapidly. At a distance of 4 m we sacrificed approximately 4-dB noise reduction.

2) Coherent and Non-Coherent Noise Field: In this section, we study the amount of noise reduction in a coherent plus non-coherent noise field with different input SNRs. The input SNR (iSNR$_{\rm nc}$) of the non-coherent noise is 20 dB. The distance between the source and the first microphone was set to $D = 2$ m. In this experiment, four microphones were used, and the reverberation time was set to $T_{60} = 0.3$ s. In Fig. 7(a), the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$ and for input SNR (iSNR$_{\rm c}$) of the coherent noise source that ranges from -5 dB to 30 dB. In Fig. 7(b), the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$ and input SNRs of -5, 0, and 30 dB. We observe the tradeoff between speech dereverberation and noise reduction as before. In addition, we see that the noise reduction in a coherent noise field is much larger than the noise reduction in a non-coherent noise field.

3) Diffuse Noise Field: In this section, we study the amount of noise reduction in a diffuse noise field. To ensure that the noise covariance matrix is full-rank we added a non-coherent noise field with an input SNR (iSNR$_{\rm nc}$) of 30 dB. The input SNR of the diffuse noise field (iSNR$_{\rm dn}$) was 0 dB. The diffuse noise signals were generated using the method described in [24]. The distance between the source and the first microphone was set to $D = 4$ m. In this experiment, four microphones were used, and the reverberation time was set to $T_{60} = 0.3$ s. In Fig. 8, the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$. In Fig. 8(b) the normalized local noise-reduction factor is shown for $\alpha = \{0, 1\}$. We observe the tradeoff between speech dereverberation and noise reduction as before. For this specific setup, the normalized local noise-reduction factor at low frequencies is lower than the normalized local noise-reduction factor at high frequencies. Locally, we see that for most frequencies we achieve higher noise-reduction when we desire only noise-reduction. However, for some frequencies the local noise-reduction factor is slightly higher when we desire complete dereverberation and noise reduction. This clearly demonstrates that we cannot guarantee the tradeoff between speech dereverberation and noise reduction locally.

VII. Conclusion

In this paper, we studied the MVDR beamformer in room acoustics. The tradeoff between speech dereverberation and noise reduction was analyzed. The results of the theoretical performance analysis are supported by the performance evaluation. The results indicate that there is a tradeoff between the achievable noise reduction and speech dereverberation. The amount of noise reduction that is sacrificed when complete dereverberation is required depends on the direct-to-reverberation ratio of the acoustic impulse response between the source and the reference microphone, and the desired response. The performance evaluation supports the theoretical analysis and demonstrates the tradeoff between speech dereverberation and noise reduction. When desiring both speech dereverberation and noise reduction the results also demonstrate that the amount of noise reduction that is sacrificed decreases when the number of microphones increases.

Appendix A
Proof of Property 5.1

Proof: Let us first evaluate the local SPCC $|\rho[X_1(j\omega), Y_1(j\omega)]|^2$ [using (2) and (18)] and $|\rho[\mathbf{h}_{\rm MVDR}^H(j\omega)\mathbf{x}(j\omega), \mathbf{h}_{\rm MVDR}^H(j\omega)\mathbf{y}(j\omega)]|^2$ [using (3) and (20)]

$$\begin{aligned}
&|\rho[X_1(j\omega), Y_1(j\omega)]|^2 \\
&\quad = \frac{|G_1(j\omega)|^2\,\phi_s(j\omega)}{|G_1(j\omega)|^2\,\phi_s(j\omega) + \phi_{v_1}(j\omega)} \\
&\quad = \frac{\text{iSNR}\,[Q(j\omega)]}{\frac{|Q(j\omega)|^2}{|G_1(j\omega)|^2} + \text{iSNR}[Q(j\omega)]},
\end{aligned} \tag{56}$$

$$\begin{aligned}
&\left|\rho\left[\mathbf{h}_{\rm MVDR}^H(j\omega)\mathbf{x}(j\omega), \mathbf{h}_{\rm MVDR}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2 \\
&\quad = \frac{\text{oSNR}[\mathbf{h}_{\rm MVDR}(j\omega)]}{1 + \text{oSNR}[\mathbf{h}_{\rm MVDR}(j\omega)]}.
\end{aligned} \tag{57}$$

In addition, we evaluate the local SPCC or MSCF between $Y_1(j\omega)$ and $\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)$

$$\begin{aligned}&\left|\rho\left[Y_1(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2\\&=\frac{|\mathbf{u}^T\mathbf{\Phi}_y(j\omega)\mathbf{h}_{\mathrm{MVDR}}(j\omega)|^2}{\phi_{y_1}(j\omega)\cdot\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{\Phi}_y(j\omega)\mathbf{h}_{\mathrm{MVDR}}(j\omega)}\\&=\frac{\phi_{x_1}(j\omega)}{\phi_{y_1}(j\omega)}\cdot\frac{C(j\omega)\phi_{x_1}(j\omega)}{\mathbf{u}^T\mathbf{\Phi}_x(j\omega)\mathbf{h}_{\mathrm{MVDR}}(j\omega)}\\&=\frac{|\rho[X_1(j\omega),Y_1(j\omega)]|^2}{\left|\rho\left[X_1(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2}.\end{aligned}\tag{58}$$

From (58) and the fact that $|\rho[A(j\omega),B(j\omega)]|^2\leq 1$ we have

$$\begin{aligned}&|\rho[X_1(j\omega),Y_1(j\omega)]|^2\\&\quad=\left|\rho\left[Y_1(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2\\&\qquad\cdot\left|\rho\left[X_1(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2\\&\quad\leq\left|\rho\left[X_1(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2.\end{aligned}\tag{59}$$

In addition, it can be shown that

$$\begin{aligned}&\left|\rho\left[X_1(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2\\&\quad=\left|\rho\left[X_1(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{x}(j\omega)\right]\right|^2\\&\qquad\times\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{x}(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2\\&\quad\leq\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{x}(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2.\end{aligned}\tag{60}$$

From (59) and (60) we know that

$$\begin{aligned}&|\rho[X_1(j\omega),Y_1(j\omega)]|^2\\&\quad\leq\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{x}(j\omega),\mathbf{h}_{\mathrm{MVDR}}^H(j\omega)\mathbf{y}(j\omega)\right]\right|^2.\end{aligned}\tag{61}$$

Hence, by substituting (56) and (57) in (61) we obtain

$$\frac{\mathrm{iSNR}[Q(j\omega)]}{\frac{|Q(j\omega)|^2}{|G_1(j\omega)|^2}+\mathrm{iSNR}[Q(j\omega)]}\leq\frac{\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]}{1+\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]}.\tag{62}$$

As a result

$$\begin{aligned}&|Q(j\omega)|^2\cdot\mathrm{oSNR}[\mathbf{h}_{\mathrm{MVDR}}(j\omega)]\\&\qquad\geq|G_1(j\omega)|^2\cdot\mathrm{iSNR}[Q(j\omega)],\quad\forall\omega\end{aligned}\tag{63}$$

which is equal to (33). ■

REFERENCES

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

[2] S. Gannot and I. Cohen, , J. Benesty, M. M. Sondhi, and Y. Huang, Eds., "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*. New York: Springer-Verlag, 2007, ch. 48, book.

[3] J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[4] S. Darlington, "Linear least-squares smoothing and prediction with applications," *Bell Syst. Tech. J.*, vol. 37, pp. 1121–1194, 1952.

[5] M. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-31, no. 6, pp. 1378–1393, Dec. 1983.

[6] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug.. 1972.

[7] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. ASSP-34, no. 6, pp. 1391–1400, Dec. 1986.

[8] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[9] 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. 2002.

[10] S. Affes and Y. Grenier, "A source subspace tracking array of microphones for double talk situations," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Munich, Germany, Apr. 1997, pp. 269–272.

[11] 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.

[12] 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, no. 3, pp. 1053–1065, Mar. 2007.

[13] Y. Huang, J. Benesty, and J. Chen, , J. Benesty, M. M. Sondhi, and Y. Huang, Eds., "Adaptive blind multichannel identification," in *Springer Handbook of Speech Processing*. New York: Springer-Verlag, 2007, ch. 13.

[14] S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP J. Appl. Signal Process.*, vol. 2003, no. 11, pp. 1074–1090, Oct. 2003.

[15] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 4, pp. 487–503, Jul. 2005.

[16] W. Herbordt, *Sound Capture for Human/Machine Interfaces—Practical Aspects of Microphone Array Signal Processing*, ser. Lecture Notes in Control and Information Sciences. Heidelberg, Germany: Springer, 2005, vol. 315.

[17] , M. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.

[18] H. L. V. Trees, *Detection, Estimation, and Modulation Theory*. New York: Wiley, Apr. 2002, vol. IV, Optimum Array Processing.

[19] J. Benesty, J. Chen, and Y. Huang, "On the importance of the Pearson correlation coefficient in noise reduction," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 4, pp. 757–765, May 2008.

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

[21] E. E. Jan and J. Flanagan, "Sound capture from spatial volumes: Matched-filter processing of microphone arrays having randomly-distributed sensors," in *Proc. IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, Atlanta, Georgia, May 1996, pp. 917–920.

[22] G. Reuven, S. Gannot, and I. Cohen, "Performance analysis of the dual source transfer-function generalized sidelobe canceller," *Speech Commun.*, vol. 49, pp. 602–622, Jul.–Aug. 2007.

[23] G. W. Elko, , M. Brandstein and D. Ward, Eds., "Spatial coherence functions," in *Microphone Arrays: Signal Processing Techniques and Applications*. New York: Springer-Verlag, 2001, ch. 4, pp. 61–85.

[24] E. Habets and S. Gannot, "Generating sensor signals in isotropic noise fields," *J. Acoust. Soc. Amer.*, vol. 122, no. 6, pp. 3464–3470, Dec. 2007.

[25] P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *J. Acoust. Soc. Amer.*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.

Emanuël A. P. Habets (S'02–M'07) received the B.Sc. degree in electrical engineering from the Hogeschool, Limburg, The Netherlands, in 1999 and the M.Sc. and Ph.D. degrees in electrical engineering from the Technische Universiteit Eindhoven, Eindhoven, The Netherlands, in 2002 and 2007, respectively.

In February 2006, he was a Guest Researcher at Bar-Ilan University, Ramat-Gan, Israel. From March 2007 to February 2009, he was a Postdoctoral Researcher at the Technion—Israel Institute of Technology and, Bar-Ilan University. Since February 2009, he has been a Research Associate at Imperial College London, London, U.K. His research interest include statistical signal processing and speech enhancement using either single or multiple microphones with applications in acoustic communication systems. His main research interest is speech dereverberation.

Dr. Habets was a member of the organization committee of the Ninth International Workshop on Acoustic Echo and Noise Control (IWAENC) in Eindhoven, The Netherlands, in 2005.

Jacob Benesty (SM'04) was born in 1963. He received the M.S. degree in microwaves from Pierre and Marie Curie University, Paris, France, in 1987, and the Ph.D. degree in control and signal processing from Orsay University, Paris, in April 1991.

During the Ph.D. degree (from November 1989 to April 1991), he worked on adaptive filters and fast algorithms at the Centre National d'Etudes des Telecommunications (CNET), Paris. From January 1994 to July 1995, he was with Telecom Paris University working 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. In May 2003, he joined INRS-EMT, University of Quebec, Montreal, QC, Canada, as a Professor. His research interests are in signal processing, acoustic signal processing, and multimedia communications. He coauthored the books *Noise Reduction in Speech Processing* (Springer-Verlag, 2009), *Microphone Array Signal Processing* (Springer-Verlag, 2008), *Acoustic MIMO Signal Processing* (Springer-Verlag, 2006), and *Advances in Network and Acoustic Echo Cancellation* (Springer-Verlag, 2001). He is the Editor-in-Chief of the reference *Springer Handbook of Speech Processing* (Springer-Verlag, Berlin, 2007). He is also a coeditor/coauthor of the books "Speech Enhancement" (Springer-Verlag, 2005), *Audio Signal Processing for Next Generation Multimedia Communication Systems* (Kluwer, 2004), *Adaptive Signal Processing: Applications to Real-World Problems* (Springer-Verlag, 2003), and *Acoustic Signal Processing for Telecommunication* (Kluwer, 2000).

Dr. Benesty received the 2001 and 2008 Best Paper Awards from the IEEE Signal Processing Society. He was a member of the editorial board of the *EURASIP Journal on Applied Signal Processing*, a member of the IEEE Audio and Electroacoustics Technical Committee, and the Co-Chair of the 1999 International Workshop on Acoustic Echo and Noise Control (IWAENC). He is the general Co-Chair of the 2009 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA).

Israel Cohen (M'01–SM'03) received the B.Sc. (*summa cum laude*), M.Sc., and Ph.D. degrees in electrical engineering from the Technion—Israel Institute of Technology, Haifa, in 1990, 1993, and 1998, respectively.

From 1990 to 1998, he was a Research Scientist with RAFAEL Research Laboratories, Haifa, Israel Ministry of Defense. From 1998 to 2001, he was a Postdoctoral Research Associate with the Computer Science Department, Yale University, New Haven, CT. In 2001, he joined the Electrical Engineering Department of the Technion, where he is currently an Associate Professor. His research interests are statistical signal processing, analysis and modeling of acoustic signals, speech enhancement, noise estimation, microphone arrays, source localization, blind source separation, system identification and adaptive filtering. He served as a Guest Editor of a special issue of the *EURASIP Journal on Advances in Signal Processing* on advances in multimicrophone speech processing and a special issue of the *EURASIP Speech Communication Journal* on speech enhancement. He is a coeditor of the Multichannel Speech Processing section of the *Springer Handbook of Speech Processing* (Springer, 2007).

Dr. Cohen received in 2005 and 2006 the Technion Excellent Lecturer awards. He served as an Associate Editor of the IEEE TRANSACTIONS ON AUDIO, SPEECH, AND LANGUAGE PROCESSING and IEEE SIGNAL PROCESSING LETTERS. He served as a Co-Chair of the 2010 International Workshop on Acoustic Echo and Noise Control.

Sharon Gannot (S'92–M'01–SM'06) received the B.Sc. degree (*summa cum laude*) from the Technion—Israel Institute of Technology, Haifa, in 1986 and the M.Sc. (*cum laude*) and Ph.D. degrees from Tel-Aviv University, Israel, in 1995 and 2000, respectively, all in electrical engineering.

In 2001, he held a postdoctoral position at the Department of Electrical Engineering (ESAT-SCD) at K.U. Leuven, Leuven, Belgium. From 2002 to 2003, he held a research and teaching position at the Faculty of Electrical Engineering, Technion. Currently, he is a Senior Lecturer at the School of Engineering, Bar-Ilan University, Israel. He is an Associate Editor of the *EURASIP Journal of Applied Signal Processing*, an Editor of two special issues on multi-microphone speech processing of the same journal, and a Guest Editor of *ELSEVIER Speech Communication Journal*.

Dr. Gannot is a reviewer of many IEEE journals and conferences. Dr. Gannot has been a member of the Technical and Steering committee of the International Workshop on Acoustic Echo and Noise Control (IWAENC) since 2005 and general Co-Chair of IWAENC 2010 to be held at Tel-Aviv, Israel. His research interests include parameter estimation, statistical signal processing and speech processing using either single- or multi-microphone arrays.

Jacek Dmochowski received the B.Eng. degree (with high distinction) in communications engineering, the M.A.S. degree in electrical engineering from Carleton University, Ottawa, ON, Canada, and the Ph.D. degree in telecommunications (granted "exceptionnelle") from the University of Quebec—INRS-EMT, in 2003, 2005, and 2008 respectively.

Dr. Dmochowski is currently a Postdoctoral Fellow at the Department of Biomedical Engineering, City College of New York, City University of New York, and is the recipient of the National Sciences and Engineering Research Council (NSERC) of Canada Post Doctoral Fellowship (2008–2010). His research interests lie in the area of multichannel statistical signal processing and include machine learning of neural signals, decoding of brain state, and neuronal modeling.

View publication stats