Chapter 9
The MVDR Beamformer for Speech Enhancement

Emanuël A. P. Habets, Jacob Benesty, Sharon Gannot, and Israel Cohen

Abstract [1]The minimum variance distortionless response (MVDR) beamformer is widely studied in the area of speech enhancement and can be used for both speech dereverberation and noise reduction. This chapter summarizes some new insights into the MVDR beamformer. Specifically, the local and global behaviors of the MVDR beamformer are analyzed, different forms of the MVDR beamformer and relations between the MVDR and other optimal beamformers are discussed. In addition, the tradeoff between dereverberation and noise reduction is analyzed. This analysis is done for a mixture of coherent and non-coherent noise fields and entirely non-coherent 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 demonstrates the tradeoff between dereverberation and noise reduction.

Emanuël A. P. Habets
Imperial College London, UK, e-mail: e.habets@imperial.ac.uk

Jacob Benesty
INRS-EMT, QC, Canada, e-mail: benesty@emt.inrs.ca

Sharon Gannot
Bar-Ilan University, Israel, e-mail: gannot@eng.biu.ac.il

Israel Cohen
Technion–Israel Institute of Technology, Israel, e-mail: icohen@ee.technion.ac.il

[1] This work was supported by the Israel Science Foundation under Grant 1085/05.

I. Cohen et al. (Eds.): Speech Processing in Modern Communication, STSP 3, pp. 225–254.
springerlink.com

9.1 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].

For single-channel noise reduction, the Wiener filter can be considered as one of the most fundamental approaches (see for example [3] and the references therein). The Wiener filter produces a minimum mean-squared error (MMSE) estimate of the desired speech component received by the microphone. Doclo and Moonen [4], proposed a multichannel Wiener Filter (MWF) technique that produces an MMSE estimate of the desired speech component in one of the microphone signals. In [5], the optimization criterion of the MWF was modified to take the allowable speech distortion into account, resulting in the speech-distortion-weighted MWF (SDW-MWF). Another interesting solution is provided by the minimum variance distortionless response (MVDR) beamformer, also known as Capon beamformer [6], which 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 [7]. Several researchers developed beamformers in which additional linear constraints were imposed (e.g., Er and Cantoni [8]). These beamformers are known as linear constraint minimum variance (LCMV) beamformers, of which the MVDR beamformer is a special case. In [9], Frost proposed an adaptive scheme of the MVDR beamformer, which is based on a constrained least-mean-square (LMS) type adaptation. Kaneda et al. [10] 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 [11] proposed the generalized sidelobe canceler (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 [12] 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 [13] and later by Gannot et al. [14]. The frequency-domain version in [14], which takes into account the reverberant

nature of the enclosure, was termed the transfer-function generalized sidelobe canceler (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 [15]. 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 [16] and the references therein and in [17]. Blind estimation of the ATFs is however beyond the scope of this chapter in which we assume that the ATFs between the source and the sensors are known. In earlier works [15], it was observed that there is a tradeoff between the amount of speech dereverberation and noise reduction. Only recently this tradeoff was analyzed by Habets et al. in [18].

In this chapter we study the MVDR beamformer in room acoustics and with broadband signals. First, we analyze the local and global behaviors [1] of the MVDR beamformer. Secondly, we derive several different forms of the MVDR filter and discuss the relations between the MVDR beamformer and other optimal beamformers. Finally, we analyze the tradeoff between noise and reverberation reduction. The local and global behaviors, as well as the tradeoff, are analyzed for different noise fields, viz. a mixture of coherent and non-coherent noise fields and entirely non-coherent noise fields.

The chapter is organized as follows: in Section 9.2 the array model is formulated and the notation used in this chapter is introduced. In Section 9.3 we start by formulating the SDW-MWF in the frequency domain. We then show that the MWF as well as the MVDR filter are special cases of the SDW-MWF. In Section 9.4 we define different performance measures that will be used in our analysis. In Section 9.5 we analyze the performance of the MVDR beamformer. The performance evaluation that demonstrates the tradeoff between reverberation and noise reduction is presented in Section 9.6. Finally, conclusions are provided in Section 9.7.

9.2 Problem Formulation

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 [19, 1]

$$\begin{aligned} y_n(k) &= g_n * s(k) + v_n(k) \\ &= x_n(k) + v_n(k),\ n = 1, 2, \ldots, N, \end{aligned} \tag{9.1}$$

where k is the discrete-time index, 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. In this chapter, without loss of generality, we consider the first microphone ($n = 1$) as the reference microphone. Our main objective 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 [1].

In the frequency domain, (9.1) can be rewritten as

$$\begin{aligned} Y_n(\omega) &= G_n(\omega)S(\omega) + V_n(\omega) \\ &= X_n(\omega) + V_n(\omega), \ n = 1, 2, \ldots, N, \end{aligned} \tag{9.2}$$

where $Y_n(\omega)$, $G_n(\omega)$, $S(\omega)$, $X_n(\omega) = G_n(\omega)S(\omega)$, and $V_n(\omega)$ 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 ω ($-\pi < \omega \leq \pi$) and j is the imaginary unit ($j^2 = -1$). We recall that the DTFT and the inverse transform [20] are

$$A(\omega) = \sum_{k=-\infty}^{\infty} a(k)e^{-j\omega k}, \tag{9.3}$$

$$a(k) = \frac{1}{2\pi}\int_{-\pi}^{\pi} A(\omega)e^{j\omega k}d\omega. \tag{9.4}$$

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

$$\begin{aligned} \mathbf{y}(\omega) &= \mathbf{g}(\omega)S(\omega) + \mathbf{v}(\omega) \\ &= \mathbf{x}(\omega) + \mathbf{v}(\omega), \end{aligned} \tag{9.5}$$

where

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

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

$$\begin{aligned}\phi_{y_n y_n}(\omega) &= \phi_{x_n x_n}(\omega) + \phi_{v_n v_n}(\omega) \\ &= |G_n(\omega)|^2 \phi_{ss}(\omega) + \phi_{v_n v_n}(\omega),\ n = 1, 2, \ldots, N,\end{aligned} \tag{9.6}$$

where $\phi_{y_n y_n}(\omega)$, $\phi_{x_n x_n}(\omega)$, $\phi_{ss}(\omega)$, and $\phi_{v_n v_n}(\omega)$ 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:

$$\begin{aligned}Z(\omega) &= \mathbf{h}^H(\omega)\mathbf{y}(\omega) \\ &= \mathbf{h}^H(\omega)\left[\mathbf{g}(\omega)S(\omega) + \mathbf{v}(\omega)\right],\end{aligned} \tag{9.7}$$

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

$$\mathbf{h}(\omega) = \left[\, H_1(\omega)\ H_2(\omega)\ \cdots\ H_N(\omega) \,\right]^T$$

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

$$\phi_{zz}(\omega) = \mathbf{h}^H(\omega)\Phi_{xx}(\omega)\mathbf{h}(\omega) + \mathbf{h}^H(\omega)\Phi_{vv}(\omega)\mathbf{h}(\omega), \tag{9.8}$$

where

$$\begin{aligned}\Phi_{xx}(\omega) &= E\left[\mathbf{x}(\omega)\mathbf{x}^H(\omega)\right] \\ &= \phi_{ss}(\omega)\mathbf{g}(\omega)\mathbf{g}^H(\omega)\end{aligned} \tag{9.9}$$

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

$$\Phi_{vv}(\omega) = E\left[\mathbf{v}(\omega)\mathbf{v}^H(\omega)\right] \tag{9.10}$$

is the PSD matrix of the noise field. In the sequel we assume that the noise is not fully coherent at the microphones so that $\Phi_{vv}(\omega)$ is a full-rank matrix and its inverse exists.

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

the definitions of performance measures and in the derivation of the various beamformers.

9.3 From Speech Distortion Weighted Multichannel Wiener Filter to Minimum Variance Distortionless Response Filter

In this section, we first formulate the SDW-MWF in the context of room acoustics. We then focus on a special case of the SDW-MWF, namely the celebrated MVDR beamformer proposed by Capon [6]. It is then shown that the SDW-MWF can be decomposed into an MVDR beamfomer and a speech distortion weighted single-channel Wiener filter. Finally, we show that the MVDR beamformer and the maximum SNR beamfomer are equivalent.

9.3.1 Speech Distortion Weighted Multichannel Wiener Filter

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

$$\begin{aligned}\mathcal{E}(\omega) &= Z(\omega) - Q(\omega)S(\omega) \\ &= \underbrace{\left[\mathbf{h}^H(\omega)\mathbf{g}(\omega) - Q(\omega)\right] S(\omega)}_{\mathcal{E}_{\tilde{s}}(\omega)} + \underbrace{\mathbf{h}^H(\omega)\mathbf{v}(\omega)}_{\mathcal{E}_v(\omega)},\end{aligned} \tag{9.11}$$

where $\tilde{S}(\omega) = Q(\omega)\, S(\omega)$. The first term $\mathcal{E}_{\tilde{s}}(\omega)$ denotes the residual desired signal at the output of the beamformer and the second term $\mathcal{E}_v(\omega)$ denotes the residual noise signal at the output of the beamformer. The mean-squared error (MSE) is given by

$$\begin{aligned}J\left[\mathbf{h}(\omega)\right] &= E\left[\left|\mathcal{E}(\omega)\right|^2\right] \\ &= E\left[\left|\mathcal{E}_{\tilde{s}}(\omega)\right|^2\right] + E\left[\left|\mathcal{E}_v(\omega)\right|^2\right] \\ &= \left|\mathbf{h}^H(\omega)\mathbf{g}(\omega) - Q(\omega)\right|^2 \phi_{ss}(\omega) + \mathbf{h}^H(\omega)\Phi_{vv}(\omega)\mathbf{h}(\omega).\end{aligned} \tag{9.12}$$

The objective of the MWF is to provide an MMSE estimate of either the clean speech source signal, the (reverberant) speech component in one of the microphone signals, or a reference signal. Therefore, the MWF inevitably introduces some speech distortion. To control the tradeoff between speech distortion and noise reduction the SDW-MWF was proposed [5, 21]. The

objective of the SDW-MWF can be described as follows[2]

$$\underset{\mathbf{h}(\omega)}{\operatorname{argmin}} E\left[|\mathcal{E}_v(\omega)|^2\right] \quad \text{subject to} \quad E\left[|\mathcal{E}_{\tilde{s}}(\omega)|^2\right] \leq \sigma^2(\omega), \tag{9.13}$$

where $\sigma^2(\omega)$ defines the maximum local power of the residual desired signal. Since the maximum local power of the residual desired signal is upper-bounded by $|Q(\omega)|^2 \phi_{ss}(\omega)$ we have $0 \leq \sigma^2(\omega) \leq |Q(\omega)|^2 \phi_{ss}(\omega)$. The solution of (9.13) can be found using the Karush-Kuhn-Tucker necessary conditions for constrained minimization [22]. Specifically, $\mathbf{h}(\omega)$ is a feasible point if it satisfies the gradient equation of the Lagrangian

$$\mathcal{L}[\mathbf{h}(\omega), \lambda(\omega)] = E\left[|\mathcal{E}_v(\omega)|^2\right] + \lambda(\omega)\left(E\left[|\mathcal{E}_{\tilde{s}}(\omega)|^2\right] - \sigma^2(\omega)\right), \tag{9.14}$$

where $\lambda(\omega)$ denotes the Lagrange multiplier for angular frequency ω and

$$\lambda(\omega)\left(E\left[|\mathcal{E}_{\tilde{s}}(\omega)|^2\right] - \sigma^2(\omega)\right) = 0, \quad \lambda(\omega) \geq 0 \text{ and } \omega \in (-\pi, \pi]. \tag{9.15}$$

The SDW-MWF can now be obtained by setting the derivative of (9.14) with respect to $\mathbf{h}(\omega)$ to zero:

$$\mathbf{h}_{\mathrm{SDW-MWF}}(\omega, \lambda) = Q^*(\omega)\left[\Phi_{xx}(\omega) + \lambda^{-1}(\omega)\Phi_{vv}(\omega)\right]^{-1}\phi_{ss}(\omega)\mathbf{g}(\omega), \tag{9.16}$$

where superscript * denotes complex conjugation. Using the Woodbury's identity (also known as the matrix inversion lemma) we can write (9.16) as

$$\mathbf{h}_{\mathrm{SDW-MWF}}(\omega, \lambda) = Q^*(\omega)\frac{\phi_{ss}(\omega)\Phi_{vv}^{-1}(\omega)\mathbf{g}(\omega)}{\lambda^{-1}(\omega) + \phi_{ss}(\omega)\mathbf{g}^H(\omega)\Phi_{vv}^{-1}(\omega)\mathbf{g}(\omega)}. \tag{9.17}$$

In order to satisfy (9.15) we require that

$$\sigma^2(\omega) = E\left[|\mathcal{E}_{\tilde{s}}(\omega)|^2\right]. \tag{9.18}$$

Therefore, the Lagrange multiplier $\lambda(\omega)$ must satisfy

[2] The employed optimization problem differs from the one used in [5, 21]. However, it should be noted that the solutions are mathematically equivalent. The advantage of the employed optimization problem is that it is directly related to the MVDR beamformer as will be shown in the following section.

$$\begin{aligned}\sigma^2(\omega) &= \left|\mathbf{h}_{\mathrm{SDW-MWF}}^H(\omega,\lambda)\mathbf{g}(\omega) - Q(\omega)\right|^2 \phi_{ss}(\omega) \\ &= \mathbf{h}_{\mathrm{SDW-MWF}}^H(\omega,\lambda)\boldsymbol{\Phi}_{xx}(\omega)\mathbf{h}_{\mathrm{SDW-MWF}}(\omega,\lambda) \\ &\quad - \mathbf{h}_{\mathrm{SDW-MWF}}^H(\omega,\lambda)\mathbf{g}(\omega)Q^*(\omega)\phi_{ss}(\omega) \\ &\quad - \mathbf{g}^H(\omega)\mathbf{h}_{\mathrm{SDW-MWF}}(\omega,\lambda)Q(\omega)\phi_{ss}(\omega) \\ &\quad + |Q(\omega)|^2 \phi_{ss}(\omega).\end{aligned} \tag{9.19}$$

Using (9.19), it is possible to find the Lagrange multiplier that results in a specific maximum local power of the residual desired signal. It can be shown that λ monotonically increases when σ^2 decreases. When $\sigma^2(\omega) = |Q(\omega)|^2 \phi_{ss}(\omega)$ we are stating that we allow maximum speech distortion. In order to satisfy (9.19), $\mathbf{h}_{\mathrm{SDW-MWF}}(\omega,\lambda)$ should be equal to $[\,0\;0\;\cdots\;0\,]^T$, which is obtain when $\lambda(\omega)$ approaches zero. Consequently, we obtain maximum noise reduction and maximum speech distortion. Another interesting solution is obtained when $\lambda(\omega) = 1$, in this case $\mathbf{h}_{\mathrm{SDW-MWF}}(\omega,\lambda)$ is equal to the non-causal multichannel Wiener filter.

For the particular case, $Q(\omega) = G_1(\omega)$, where we only want to reduce the level of the noise (no dereverberation at all), we can eliminate the explicit dependence of (9.17) on the acoustic transfer functions. Specifically, by using the fact that $\phi_{ss}(\omega)\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)$ is equal to $\mathrm{tr}\left[\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{xx}(\omega)\right]$ we obtain the following forms:

$$\begin{aligned}\mathbf{h}_{\mathrm{SDW-MWF}}(\omega,\lambda) &= \frac{\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{xx}(\omega)}{\lambda^{-1}(\omega) + \mathrm{tr}\left[\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{xx}(\omega)\right]}\mathbf{u} \\ &= \frac{\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{yy}(\omega) - \mathbf{I}}{\lambda^{-1}(\omega) + \mathrm{tr}\left[\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{yy}(\omega)\right] - N}\mathbf{u},\end{aligned} \tag{9.20}$$

where $\mathrm{tr}(\cdot)$ denotes the trace of a matrix, $\mathbf{I}$ is the $N \times N$ identity matrix,

$$\boldsymbol{\Phi}_{yy}(\omega) = E\left[\mathbf{y}(\omega)\mathbf{y}^H(\omega)\right] \tag{9.21}$$

is the PSD matrix of the microphone signals, and $\mathbf{u} = [\,1\;0\;\cdots\;0\;0\,]^T$ is a vector of length N.

9.3.2 Minimum Variance Distortionless Response Filter

The MVDR filter can be found by minimizing the local power of the residual desired signal at the output of the beamformer. This can be achieved by setting the maximum local power of the residual desired signal $\sigma^2(\omega)$ in (9.13) equal to zero. We then obtain the following optimalization problem

$$\mathbf{h}_{\mathrm{MVDR}}(\omega) = \underset{\mathbf{h}(\omega)}{\operatorname{argmin}} E\left[|\mathcal{E}_v(\omega)|^2\right] \quad \text{subject to} \quad E\left[|\mathcal{E}_{\tilde{s}}(\omega)|^2\right] = 0, \tag{9.22}$$

Alternatively, we can use the MSE in (9.12) to derive the MVDR filter, which is conceived by providing a fixed gain [in our case modelled by $Q(\omega)$] 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 (9.12)] to the array output. The latter optimization problem can be formulated as

$$\mathbf{h}_{\mathrm{MVDR}}(\omega) = \underset{\mathbf{h}(\omega)}{\operatorname{argmin}} E\left[|\mathcal{E}_v(\omega)|^2\right] \quad \text{subject to} \quad \mathbf{h}^H(\omega)\mathbf{g}(\omega) = Q(\omega). \tag{9.23}$$

Since $E\left[|\mathcal{E}_{\tilde{s}}(\omega)|^2\right] = \left|\mathbf{h}^H(\omega)\mathbf{g}(\omega) - Q(\omega)\right|^2 \phi_{ss}(\omega) = 0$ for $\mathbf{h}^H(\omega)\mathbf{g}(\omega) = Q(\omega)$ we obtain the same solution for both optimization problems, i.e.,

$$\mathbf{h}_{\mathrm{MVDR}}(\omega) = Q^*(\omega)\frac{\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}{\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}. \tag{9.24}$$

The MVDR filter can also be obtained from the SDW-MWF defined in (9.17) by finding the Lagrange multiplier $\lambda(\omega)$ that satisfies (9.15) for $\sigma^2(\omega) = 0$. To satisfy (9.15) we require that the local power of the residual desired signal at the output of the beamformer, $E\left[|\mathcal{E}_{\tilde{s}}(\omega)|^2\right]$, is equal to zero. From (9.19) it can be shown directly that $\sigma^2(\omega) = 0$ when $\mathbf{h}_{\mathrm{SDW-MWF}}^H(\omega, \lambda)\mathbf{g}(\omega) = Q(\omega)$. Using (9.17) the latter expression can be written as

$$Q(\omega)\frac{\phi_{ss}(\omega)\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}{\lambda^{-1}(\omega) + \phi_{ss}(\omega)\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)} = Q(\omega). \tag{9.25}$$

Hence, when $\lambda(\omega)$ goes to infinity the left and right hand sides of (9.25) are equal. Consequently, we have

$$\lim_{\lambda(\omega)\to\infty} \mathbf{h}_{\mathrm{SDW-MWF}}(\omega, \lambda) = \mathbf{h}_{\mathrm{MVDR}}(\omega). \tag{9.26}$$

We can get rid of the explicit dependence on the acoustic transfer functions $\{G_2(\omega), \ldots, G_M(\omega)\}$ of the MVDR filter (9.24) by multiplying the numerator and denominator in (9.24) by $\phi_{ss}(\omega)$ and using the fact that $\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)$ is equal to $\mathrm{tr}\left[\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)\mathbf{g}^H(\omega)\right]$ to obtain the following form [18]:

$$\mathbf{h}_{\mathrm{MVDR}}(\omega) = \frac{Q^*(\omega)}{G_1^*(\omega)}\frac{\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{xx}(\omega)}{\mathrm{tr}\left[\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{xx}(\omega)\right]}\mathbf{u}. \tag{9.27}$$

Basically, we only need $G_1(\omega)$ to achieve dereverberation and noise reduction. It should however be noted that $\mathbf{h}_{\mathrm{MVDR}}(\omega)$ is a non-causal filter.

Using the Woodbury's identity, another important form of the MVDR filter is derived [18]:

$$\mathbf{h}_{\text{MVDR}}(\omega) = C(\omega)\mathbf{\Phi}_{yy}^{-1}(\omega)\mathbf{\Phi}_{xx}(\omega)\mathbf{u}, \tag{9.28}$$

where

$$C(\omega) = \frac{Q^*(\omega)}{G_1^*(\omega)}\left\{1 + \frac{1}{\text{tr}\left[\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{xx}(\omega)\right]}\right\}. \tag{9.29}$$

For the particular case, $Q(\omega) = G_1(\omega)$, 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]:

$$\begin{aligned}\mathbf{h}_{\text{MVDR}}(\omega) &= \frac{\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{xx}(\omega)}{\text{tr}\left[\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{xx}(\omega)\right]}\mathbf{u} \\ &= \frac{\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{yy}(\omega) - \mathbf{I}}{\text{tr}\left[\mathbf{\Phi}_{vv}^{-1}(\omega)\mathbf{\Phi}_{yy}(\omega)\right] - N}\mathbf{u}.\end{aligned} \tag{9.30}$$

Hence, noise reduction can be achieved without explicitly estimating the acoustic transfer functions.

9.3.3 Decomposition of the Speech Distortion Weighted Multichannel Wiener Filter

Using (9.17) and (9.24) the SDW-MWF can be decomposed into an MVDR beamformer and a speech distortion weighted single-channel Wiener filter, i.e.,

$$\mathbf{h}_{\text{SDW}-\text{MWF}}(\omega, \lambda) = \mathbf{h}_{\text{MVDR}}(\omega) \cdot h_{\text{SDW-WF}}(\omega, \lambda), \tag{9.31}$$

where

$$\begin{aligned}h_{\text{SDW-WF}}(\omega, \lambda) &= \frac{\phi_{\tilde{s}\tilde{s}}(\omega)}{\phi_{\tilde{s}\tilde{s}}(\omega) + \lambda^{-1}(\omega)\,\mathbf{g}^H(\omega)\mathbf{\Phi}_{vv}(\omega)\mathbf{g}(\omega)} \\ &= \frac{\phi_{\tilde{s}\tilde{s}}(\omega)}{\phi_{\tilde{s}\tilde{s}}(\omega) + \lambda^{-1}(\omega)\,\mathbf{h}_{\text{MVDR}}^H(\omega)\mathbf{\Phi}_{vv}(\omega)\mathbf{h}_{\text{MVDR}}(\omega)},\end{aligned} \tag{9.32}$$

and $\phi_{\tilde{s}\tilde{s}}(\omega) = |Q(\omega)|^2\phi_{ss}(\omega)$. Indeed, for $\lambda(\omega) \to \infty$ (i.e., no speech distortion) the (single-channel) speech distortion weighted Wiener filter $h_{\text{SDW-WF}}(\omega, \lambda) = 1$ for all ω.

9.3.4 Equivalence of MVDR and Maximum SNR Beamformer

It is interesting to show the equivalence between the MVDR filter (9.24) and the maximum SNR (MSNR) beamformer [23], which is obtained from

$$\mathbf{h}_{\mathrm{MSNR}}(\omega) = \underset{\mathbf{h}(\omega)}{\operatorname{argmax}} \frac{|\mathbf{h}^H(\omega)\mathbf{g}(\omega)|^2 \phi_{ss}(\omega)}{\mathbf{h}^H(\omega)\boldsymbol{\Phi}_{vv}(\omega)\mathbf{h}(\omega)}. \tag{9.33}$$

The well-known solution to (9.33) is the (colored noise) matched filter

$$\mathbf{h}(\omega) \propto \boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega).$$

If the array response is constrained to fulfil $\mathbf{h}^H(\omega)\mathbf{g}(\omega) = Q(\omega)$ we have

$$\mathbf{h}_{\mathrm{MSNR}}(\omega) = Q^*(\omega) \frac{\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}{\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}. \tag{9.34}$$

This solution is identical to the solution of the MVDR filter (9.24).

9.4 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 the parameterized desired signal [given by $Q(\omega)S(\omega)$] and the noise signal received by the first microphone, i.e.,

$$\mathrm{iSNR}\left[Q(\omega)\right] = \frac{|Q(\omega)|^2 \phi_{ss}(\omega)}{\phi_{v_1 v_1}(\omega)},\ \omega \in (-\pi, \pi], \tag{9.35}$$

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

$$\mathrm{iSNR}(Q) = \frac{\int_{-\pi}^{\pi} |Q(\omega)|^2 \phi_{ss}(\omega) d\omega}{\int_{-\pi}^{\pi} \phi_{v_1 v_1}(\omega) d\omega}. \tag{9.36}$$

After applying the MVDR on the received signals, given by (9.8), the local output SNR is

$$\begin{aligned}\text{oSNR}\left[\mathbf{h}_{\text{MVDR}}(\omega)\right] &= \frac{\left|\mathbf{h}_{\text{MVDR}}^H(\omega)\mathbf{g}(\omega)\right|^2\phi_{ss}(\omega)}{\mathbf{h}_{\text{MVDR}}^H(\omega)\boldsymbol{\Phi}_{vv}(\omega)\mathbf{h}_{\text{MVDR}}(\omega)}\\ &= \frac{|Q(\omega)|^2\phi_{ss}(\omega)}{\mathbf{h}_{\text{MVDR}}^H(\omega)\boldsymbol{\Phi}_{vv}(\omega)\mathbf{h}_{\text{MVDR}}(\omega)}.\end{aligned} \tag{9.37}$$

By substituting (9.24) in (9.37) it can be shown that

$$\begin{aligned}\text{oSNR}\left[\mathbf{h}_{\text{MVDR}}(\omega)\right] &= \frac{|Q(\omega)|^2\phi_{ss}(\omega)}{|Q(\omega)|^2\frac{\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)}{\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}\boldsymbol{\Phi}_{vv}(\omega)\frac{\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}{\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)}}\\ &= \phi_{ss}(\omega)\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)\\ &= \text{tr}\left[\boldsymbol{\Phi}_{vv}^{-1}(\omega)\boldsymbol{\Phi}_{xx}(\omega)\right],\ \omega\in(-\pi,\pi].\end{aligned} \tag{9.38}$$

It is extremely important to observe that the desired response $Q(\omega)$ 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

$$\begin{aligned}\text{oSNR}\left(\mathbf{h}_{\text{MVDR}}\right) &= \frac{\int_{-\pi}^{\pi}\left|\mathbf{h}_{\text{MVDR}}^H(\omega)\mathbf{g}(\omega)\right|^2\phi_{ss}(\omega)d\omega}{\int_{-\pi}^{\pi}\mathbf{h}_{\text{MVDR}}^H(\omega)\boldsymbol{\Phi}_{vv}(\omega)\mathbf{h}_{\text{MVDR}}(\omega)d\omega}\\ &= \frac{\int_{-\pi}^{\pi}|Q(\omega)|^2\phi_{ss}(\omega)d\omega}{\int_{-\pi}^{\pi}|Q(\omega)|^2\left[\mathbf{g}^H(\omega)\boldsymbol{\Phi}_{vv}^{-1}(\omega)\mathbf{g}(\omega)\right]^{-1}d\omega}\\ &= \frac{\int_{-\pi}^{\pi}|Q(\omega)|^2\phi_{ss}(\omega)d\omega}{\int_{-\pi}^{\pi}\text{oSNR}^{-1}\left[\mathbf{h}_{\text{MVDR}}(\omega)\right]|Q(\omega)|^2\phi_{ss}(\omega)d\omega}.\end{aligned} \tag{9.39}$$

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

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:

$$\begin{aligned}\xi_{\text{nr}}\left[\mathbf{h}(\omega)\right] &= \frac{\phi_{v_1v_1}(\omega)}{\mathbf{h}^H(\omega)\boldsymbol{\Phi}_{vv}(\omega)\mathbf{h}(\omega)}\\ &= \frac{\text{oSNR}\left[\mathbf{h}(\omega)\right]}{\text{iSNR}\left[Q(\omega)\right]}\cdot\frac{|Q(\omega)|^2}{\left|\mathbf{h}^H(\omega)\mathbf{g}(\omega)\right|^2},\ \omega\in(-\pi,\pi].\end{aligned} \tag{9.40}$$

We see that $\xi_{\text{nr}}\left[\mathbf{h}(\omega)\right]$ 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 $\mathbf{h}(\omega)$. For the MVDR beamformer we have $\left|\mathbf{h}_{\text{MVDR}}^H(\omega)\mathbf{g}(\omega)\right|^2 = |Q(\omega)|^2$. Therefore we can further simplify (9.40):

$$\xi_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] = \frac{\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]}{\mathrm{iSNR}\left[Q(\omega)\right]},\ \omega \in (-\pi, \pi]. \tag{9.41}$$

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 (9.40) yields the global noise-reduction factor:

$$\begin{aligned}\xi_{\mathrm{nr}}(\mathbf{h}) &= \frac{\int_{-\pi}^{\pi} \phi_{v_1 v_1}(\omega) d\omega}{\int_{-\pi}^{\pi} \mathbf{h}^H(\omega) \mathbf{\Phi}_{vv}(\omega) \mathbf{h}(\omega) d\omega} \\ &= \frac{\mathrm{oSNR}(\mathbf{h})}{\mathrm{iSNR}(Q)} \cdot \frac{\int_{-\pi}^{\pi} |Q(\omega)|^2 \phi_{ss}(\omega) d\omega}{\int_{-\pi}^{\pi} |\mathbf{h}^H(\omega)\mathbf{g}(\omega)|^2 \phi_{ss}(\omega) d\omega}.\end{aligned} \tag{9.42}$$

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

$$\xi_{\mathrm{nr}}(\mathbf{h}_{\mathrm{MVDR}}) = \frac{\mathrm{oSNR}(\mathbf{h}_{\mathrm{MVDR}})}{\mathrm{iSNR}(Q)}. \tag{9.43}$$

9.5 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 Subsection 9.5.1 we will discuss this issue and propose a viable comparison method. In Subsections 9.5.2 and 9.5.3, we analyze the local and global behaviors 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 Subsections 9.5.4 and 9.5.5 we analyze the MVDR performance in non-coherent noise fields and mixed coherent and non-coherent noise fields, respectively.

9.5.1 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 $\mathbf{h}^H(\omega)\mathbf{g}(\omega) = G_1(\omega)$. When we

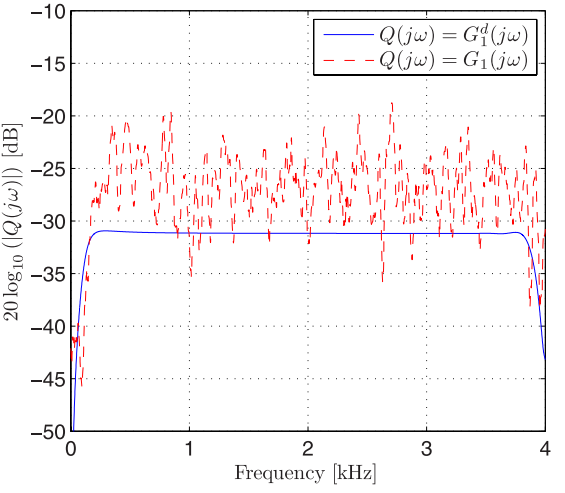


Fig. 9.1 Magnitude of the transfer functions $Q(\omega) = \{G_1(\omega), G_1^{\mathrm{d}}(\omega)\}$ (reverberation time $T_{60} = 0.5$ s, source-receiver distance $D = 2.5$ m).

desire complete dereverberation and noise reduction we can use the constraint $\mathbf{h}^H(\omega)\mathbf{g}(\omega) = G_1^{\mathrm{d}}(\omega)$, where $G_1^{\mathrm{d}}(\omega)$ denotes the transfer function of the direct path response from the source to the first microphone. In Fig. 9.1 the magnitude of the transfer functions $G_1(\omega)$ and $G_1^{\mathrm{d}}(\omega)$ are shown. The transfer function $G_1(\omega)$ was generated using the image-method [24], the distance between the source and the microphone was 2.5 m and the reverberation time was 500 ms. The transfer function $G_1^{\mathrm{d}}(\omega)$ was obtained by considering only the direct path. As expected from a physical point of view, we can see that the energy of $G_1(\omega)$ is larger than the energy of $G_1^{\mathrm{d}}(\omega)$. In addition we observe that for very few frequencies $|G_1(\omega)|^2$ is smaller than $|G_1^{\mathrm{d}}(\omega)|^2$. Evidently, the power of the desired signal $G_1^{\mathrm{d}}(\omega)S(\omega)$ is always smaller than the power of the desired signal $G_1(\omega)S(\omega)$.

Now let us first look at an illustrative example. Obviously, by choosing any constraint $Q(\omega, \gamma) = \gamma \cdot G_1^{\mathrm{d}}(\omega)$ $(\gamma > 0 \ \wedge \ \gamma \in \mathbb{R})$ we desire both noise reduction and complete dereverberation. Now let us define the MVDR filter with the constraint $Q(\omega, \gamma)$ by $\mathbf{h}_{\mathrm{MVDR}}(\omega, \gamma)$. Using (9.24) it can be shown that $\mathbf{h}_{\mathrm{MVDR}}(\omega, \gamma)$ is equal to $\gamma\, \mathbf{h}_{\mathrm{MVDR}}(\omega)$, 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 $\mathbf{h}_{\mathrm{MVDR}}(\omega)$ and $\mathbf{h}_{\mathrm{MVDR}}(\omega, \gamma)$ using (9.41) we obtain different results, i.e.,

$$\xi_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] \neq \xi_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega, \gamma)\right] \quad \text{for} \quad \gamma \neq 1. \tag{9.44}$$

This can also be explained by the fact that the local output SNRs of all MVDR beamformers $\mathbf{h}_{\mathrm{MVDR}}(\omega,\gamma)$ are equal because the local output SNR [as defined in (9.37)] is independent of γ while the local input SNR [as defined in (9.35)] 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, which can be achieved my normalizing the MVDR filter. Fundamentally, the definition of the MVDR beamformer depends on $Q(\omega)$. Therefore, the choice of different desired signals [given by $Q(\omega)S(\omega)$] is part of the (local and global) input SNR definitions. Basically we can apply any normalization provided that the power of the desired signals at the output of the beamformer is equal. However, to obtain a meaningful output power and to be consistent with earlier works, we propose to make the power of the desired signal at the output equal to the power of the signal that would be obtained when using the constraint $\mathbf{h}^H(\omega)\mathbf{g}(\omega) = G_1(\omega)$. The global normalization factor $\eta(Q,G_1)$ is therefore given by

$$\eta(Q,G_1) = \sqrt{\frac{\int_{-\pi}^{\pi} |G_1(\omega)|^2 \, \phi_{ss}(\omega)\, d\omega}{\int_{-\pi}^{\pi} |Q(\omega)|^2 \, \phi_{ss}(\omega)\, d\omega}}, \tag{9.45}$$

which can either be applied to the output signal of the beamformer or the filter $\mathbf{h}_{\mathrm{MVDR}}(\omega)$.

9.5.2 Local Analyzes

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

Property 9.1. With the MVDR filter given in (9.24), the local output SNR times $|Q(\omega)|^2$ is always greater than or equal to the local input SNR times $|G_1(\omega)|^2$, i.e.,

$$|Q(\omega)|^2 \cdot \mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] \geq |G_1(\omega)|^2 \cdot \mathrm{iSNR}\left[Q(\omega)\right],\ \forall\omega, \tag{9.46}$$

which can also be expressed using (9.35) as

$$\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] \geq \frac{|G_1(\omega)|^2 \, \phi_{ss}(\omega)}{\phi_{v_1v_1}(\omega)},\ \forall\omega. \tag{9.47}$$

Proof. See Appendix.

This property proofs that the local output SNR obtained using the MVDR filter will always be equal or larger than the ratio of the reverberant desired signal power and the noise power received by the reference microphone (in this case the first microphone).

The normalized local noise-reduction factor is defined as

$$\begin{aligned}\tilde{\xi}_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] &= \xi_{\mathrm{nr}}\left[\eta(Q,G_1)\,\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] \\ &= \frac{1}{\eta^2(Q,G_1)}\,\frac{\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]}{\mathrm{iSNR}\left[Q(\omega)\right]} \\ &= \frac{1}{\eta^2(Q,G_1)\left|Q(\omega)\right|^2}\cdot\frac{\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]\cdot\phi_{v_1v_1}(\omega)}{\phi_{ss}(\omega)} \\ &= \frac{1}{\zeta[Q(\omega),G_1(\omega)]}\cdot\frac{\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]\cdot\phi_{v_1v_1}(\omega)}{\phi_{ss}(\omega)}, \end{aligned} \tag{9.48}$$

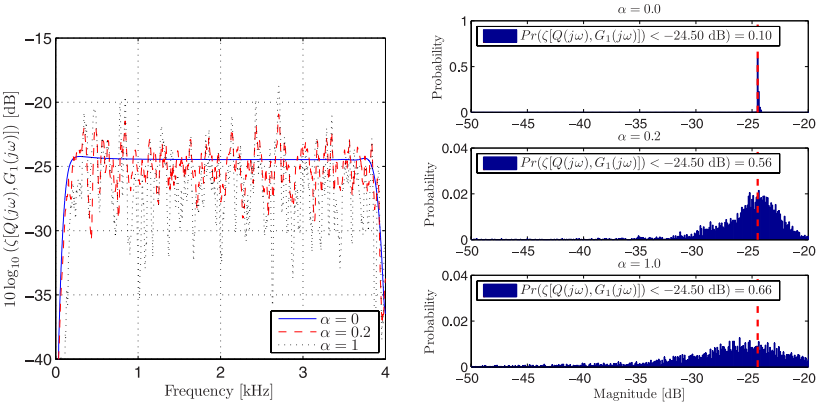
where $\zeta[Q(\omega),G_1(\omega)] = \eta^2(Q,G_1)\left|Q(\omega)\right|^2$. Indeed, for different MVDR beamformers the noise-reduction factor varies due to $\zeta[Q(\omega),G_1(\omega)]$, since the local output SNR, $\phi_{v_1v_1}(\omega)$, and $\phi_{ss}(\omega)$ do not depend on $Q(\omega)$.
Since $\zeta[Q(\omega),G_1(\omega)] = \zeta[\gamma\,Q(\omega),G_1(\omega)]$ $(\gamma > 0)$ the normalized local noise-reduction factor is independent of the global scaling factor γ.

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

$$Q(\omega,\alpha) = G_1^{\mathrm{d}}(\omega) + \alpha\,G_1^{\mathrm{r}}(\omega), \tag{9.49}$$

where the parameter $0 \leq \alpha \leq 1$ controls the direct-to-reverberation ratio (DRR) of the desired response. In Fig. 9.2(a) we plotted $\zeta[Q(\omega,\alpha),G_1(\omega)]$ for $\alpha = \{0, 0.2, 1\}$. Due to the normalization the energy of $\zeta[Q(\omega,\alpha),G_1(\omega)]$ (and therefore its mean value) does not depend on α. Locally we can see that the deviation with respect to $|G_1^{\mathrm{d}}(\omega)|^2$ increases when α increases (i.e., when the DRR decreases). In Fig. 9.2(b) we plotted the histogram of $\zeta[Q(\omega,\alpha),G_1(\omega)]$ for $\alpha = \{0, 0.2, 1\}$. First, we observe that the probability that $\zeta[Q(\omega,\alpha),G_1(\omega)]$ is smaller than its mean value decreases when α decreases (i.e., when the DRR increases). Secondly, we observe that the distribution is stretched out towards negative values when α increases. Hence, when the desired speech signal contains less reverberation it is more likely that $\zeta[Q(\omega,\alpha),G_1(\omega)]$ 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 $Q(\omega) = G_1(\omega)$.

Using Property 9.1 we deduce a lower bound for the normalized local noise-reduction factor:



(a) (b)

Fig. 9.2 a) The normalized transfer functions $\zeta[Q(\omega,\alpha),G_1(\omega)]$ with $Q(\omega,\alpha) = G_1^{\mathrm{d}}(\omega) + \alpha\, G_1^{\mathrm{r}}(\omega)$ for $\alpha = \{0, 0.2, 1\}$, b) the histograms of $10\log_{10}(\zeta[Q(\omega,\alpha),G_1(\omega)])$.

$$\tilde{\xi}_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] \geq \frac{1}{\eta^2(Q,G_1)\left|Q(\omega)\right|^2}\left|G_1(\omega)\right|^2. \tag{9.50}$$

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

$$\tilde{\xi}_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] \geq 1. \tag{9.51}$$

Expression (9.51) 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.

9.5.3 Global Analyzes

Using (9.43), (9.39), and (9.36) we deduce the normalized global noise-reduction factor:

$$\begin{aligned}\tilde{\xi}_{\mathrm{nr}}(\mathbf{h}_{\mathrm{MVDR}}) &= \xi_{\mathrm{nr}}(\eta(Q,G_1)\,\mathbf{h}_{\mathrm{MVDR}}) \\ &= \frac{1}{\eta^2(Q,G_1)}\,\frac{\mathrm{oSNR}(\mathbf{h}_{\mathrm{MVDR}})}{\mathrm{iSNR}(Q)} \\ &= \frac{1}{\eta^2(Q,G_1)}\,\frac{\int_{-\pi}^{\pi}\phi_{v_1v_1}(\omega)d\omega}{\int_{-\pi}^{\pi}\mathrm{oSNR}^{-1}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]|Q(\omega)|^2\,\phi_{ss}(\omega)d\omega} \\ &= \frac{\int_{-\pi}^{\pi}\phi_{v_1v_1}(\omega)d\omega}{\int_{-\pi}^{\pi}\mathrm{oSNR}^{-1}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]\,\zeta[Q(\omega),G_1(\omega)]\,\phi_{ss}(\omega)d\omega}. \end{aligned} \tag{9.52}$$

This normalized global noise-reduction factor behaves, with respect to $Q(\omega)$, similarly to its local counterpart. It can be verified that the normalized global noise-reduction factor for $\gamma \cdot Q(\omega)$ is independent of γ. Due to the complexity of (9.52) it is difficult to predict the exact behavior of the normalized global noise-reduction factor. From the analyzes in the previous subsection we do know that the distribution of $\zeta[Q(\omega),G_1(\omega)]$ is stretched out towards zero when the DRR decreases. Hence, for each frequency it is likely that $\zeta[Q(\omega),G_1(\omega)]$ 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 9.6.

9.5.4 Non-Coherent Noise Field

Let us assume that the noise field is non-coherent, also known as spatially white. In case the noise variance at each microphone is equal to $\sigma_{\mathrm{nc}}^2(\omega)$ the noise covariance matrix $\mathbf{\Phi}_{vv}(\omega)$ simplifies to $\sigma_{\mathrm{nc}}^2(\omega)\mathbf{I}$. In the latter case the MVDR beamformer simplifies to

$$\mathbf{h}_{\mathrm{MVDR}}(\omega) = Q^*(\omega)\frac{\mathbf{g}(\omega)}{\|\mathbf{g}(\omega)\|^2}, \tag{9.53}$$

where $\|\mathbf{g}(\omega)\|^2 = \mathbf{g}^H(\omega)\mathbf{g}(\omega)$. For $Q(\omega) = G_1(\omega)$ this is the well-known matched beamformer [25], which generalizes the delay-and-sum beamformer. The normalized local noise-reduction factor can be deduced by substituting $\sigma_{\mathrm{nc}}^2(\omega)\mathbf{I}$ in (9.48), and result in

$$\tilde{\xi}_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] = \frac{1}{\zeta[Q(\omega),G_1(\omega)]}\,\|\mathbf{g}(\omega)\|^2. \tag{9.54}$$

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

$$\tilde{\xi}_{\text{nr}}\left[\mathbf{h}_{\text{MVDR}}(\omega)\right] = \frac{\|\mathbf{g}(\omega)\|^2}{|G_1(\omega)|^2} = \left(1 + \sum_{n=2}^{N} \frac{|G_n(\omega)|^2}{|G_1(\omega)|^2}\right). \tag{9.55}$$

As we expected from (9.51), the normalized noise-reduction factor is always larger than 1 when $Q(\omega) = G_1(\omega)$. However, in other situations we cannot guarantee that there is noise reduction.

The normalized global noise-reduction factor is given by

$$\tilde{\xi}_{\text{nr}}\left(\mathbf{h}_{\text{MVDR}}\right) = \frac{1}{\eta^2(Q, G_1)} \frac{\int_{-\pi}^{\pi} \sigma_{\text{nc}}^{-2}(\omega)\ \|\mathbf{g}(\omega)\|^2\ \phi_{ss}(\omega)\, d\omega}{\int_{-\pi}^{\pi} \sigma_{\text{nc}}^{-2}(\omega)\, |Q(\omega)|^2\, \phi_{ss}(\omega)\, d\omega} = \frac{\int_{-\pi}^{\pi} \sigma_{\text{nc}}^{-2}(\omega)\ \|\mathbf{g}(\omega)\|^2\ \phi_{ss}(\omega)\, d\omega}{\int_{-\pi}^{\pi} \sigma_{\text{nc}}^{-2}(\omega)\, \zeta[Q(\omega), G_1(\omega)]\, \phi_{ss}(\omega)\, d\omega}. \tag{9.56}$$

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

$$\tilde{\xi}_{\text{nr}}\left(\mathbf{h}_{\text{MVDR}}\right) = N. \tag{9.57}$$

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 (9.30), i.e., we do not require any *a prior* knowledge of the direction of arrival.

9.5.5 Coherent plus Non-Coherent Noise Field

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

$$\Phi_{vv}(\omega) = \sigma_{\text{c}}^2(\omega)\mathbf{d}(\omega)\mathbf{d}^H(\omega) + \sigma_{\text{nc}}^2(\omega)\mathbf{I}. \tag{9.58}$$

Using Woodbury's identity the MVDR beamformer becomes

$$\mathbf{h}_{\text{MVDR}}(\omega) = Q^*(\omega) \frac{\left(\mathbf{I} - \frac{\mathbf{d}(\omega)\mathbf{d}^H(\omega)}{\frac{\sigma_{\text{nc}}^2(\omega)}{\sigma_{\text{c}}^2(\omega)} + \mathbf{d}^H(\omega)\mathbf{d}(\omega)}\right)\mathbf{g}(\omega)}{\mathbf{g}^H(\omega)\left(\mathbf{I} - \frac{\mathbf{d}(\omega)\mathbf{d}^H(\omega)}{\frac{\sigma_{\text{nc}}^2(\omega)}{\sigma_{\text{c}}^2(\omega)} + \mathbf{d}^H(\omega)\mathbf{d}(\omega)}\right)\mathbf{g}(\omega)}. \tag{9.59}$$

The normalized local noise-reduction factor is given by [18]

$$\tilde{\xi}_{\mathrm{nr}}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right] = C(\omega)\left(\mathbf{g}^H(\omega)\mathbf{g}(\omega) - \frac{|\mathbf{g}^H(\omega)\mathbf{d}(\omega)|^2}{\frac{\sigma^2_{\mathrm{nc}}(\omega)}{\sigma^2_{\mathrm{c}}(\omega)} + \mathbf{d}^H(\omega)\mathbf{d}(\omega)}\right), \tag{9.60}$$

where

$$C(\omega) = \frac{1}{\zeta[Q(\omega), G_1(\omega)]}\left(1 + \frac{\sigma^2_{\mathrm{c}}(\omega)}{\sigma^2_{\mathrm{nc}}(\omega)}\,|D_1(\omega)|^2\right). \tag{9.61}$$

The noise reduction depends on $\zeta[Q(\omega), G_1(\omega)]$, the ratio between the variance of the non-coherent and coherent, and on the inner product of $\mathbf{d}(\omega)$ and $\mathbf{g}(\omega)$ [26].

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

$$\lim_{\frac{\sigma^2_{\mathrm{nc}}(\omega)}{\sigma^2_{\mathrm{c}}(\omega)} \to 0} \mathbf{h}^H_{\mathrm{MVDR}}(\omega)\mathbf{d}(\omega)\sigma_{\mathrm{c}}(\omega) = 0.$$

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

9.6 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 scenarios. A linear microphone array was used with 2 to 8 microphones and an inter-microphone distance of 5 cm. The room size is $5 \times 4 \times 6$ m (length×width×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 [24] with some necessary modifications that ensure proper inter-microphone phase delays as proposed by Peterson [27]. 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 $Q(\omega)$. As proposed in Section 9.5.1, we control the DRR using the parameter α ($0 \leq \alpha \leq 1$). The complex scaling factor $Q(\omega, \alpha)$ is calculated using (9.49). When the desired response equals $Q(\omega, 0) = G_1^{\mathrm{d}}(\omega)$ we desire both noise reduction and complete dereverberation. However, when the desired response equals $Q(\omega, 1) = G_1(\omega)$ we desire only noise reduction.

9.6.1 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 $Q(\omega, 0) = G_1^{\mathrm{d}}(\omega)$] was iSNR = 5 dB. In this experiment 2, 4, or 8 microphones were used. In Fig. 9.3 the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$. Firstly, 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. Secondly, 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.

9.6.2 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 $Q(\omega) = G_1^{\mathrm{d}}(\omega)$] 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 $Q(\omega)$ is shown in Fig. 9.4(a). In Fig. 9.4(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

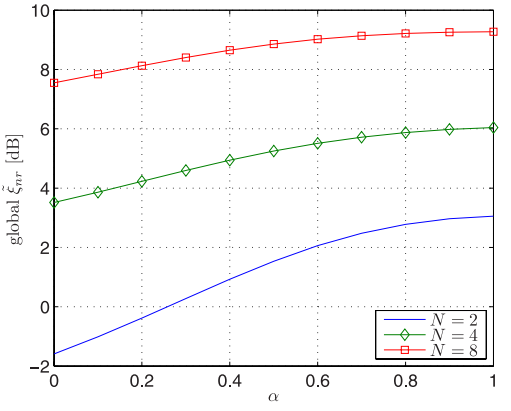
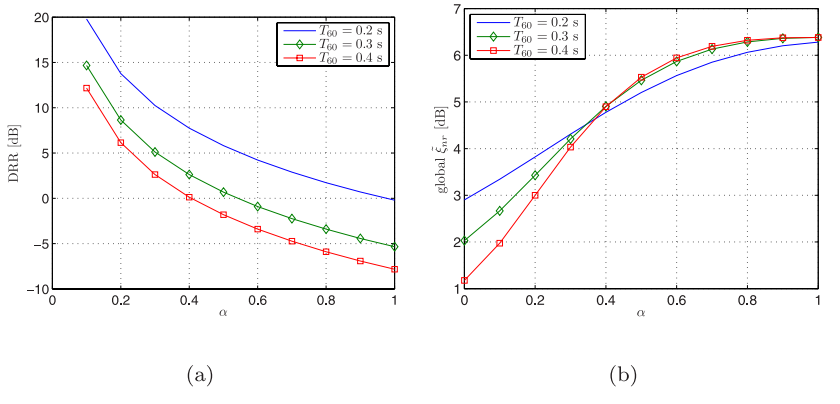


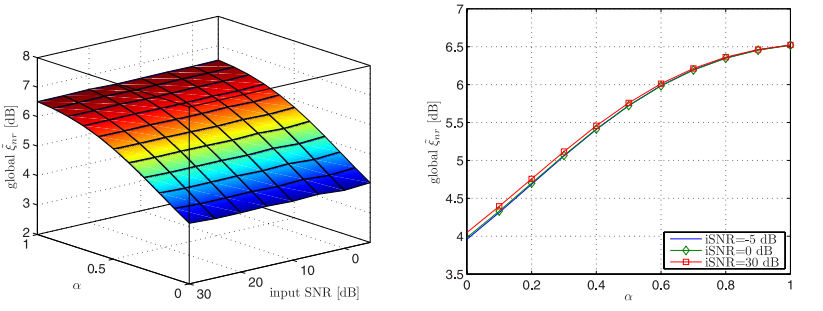
Fig. 9.3 The normalized global noise-reduction factor obtained using $N = \{2, 4, 8\}$ ($T_{60} = 0.3$, $D = 2$ m, non-coherent noise iSNR = 5 dB).



(a) (b)

Fig. 9.4 a) The DRR of $Q(\omega, \alpha)$ 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).

increase the DRR to approximately -5 dB 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.



(a) (b)

Fig. 9.5 The 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).

9.6.3 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.

9.6.3.1 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. 9.5(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. 9.5(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 (9.56), for a non-coherent noise field the normalized global noise-reduction factor is independent of the input SNR.

In Fig. 9.6, 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. When we

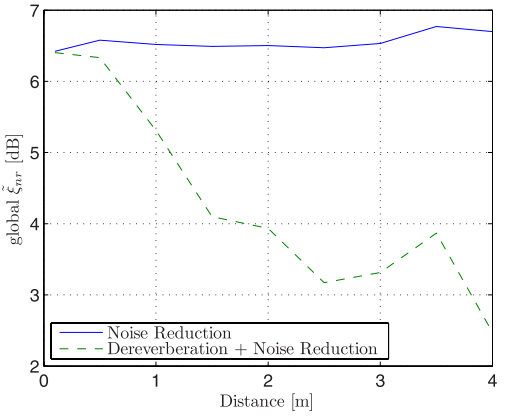
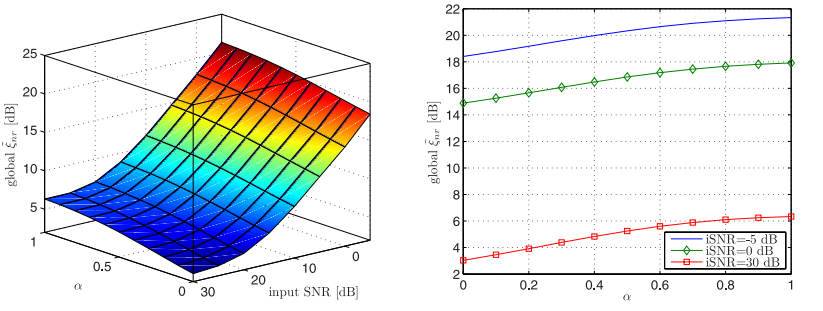


Fig. 9.6 The normalized global noise-reduction factor for one specific source trajectory obtained using $D = \{0.1, 0.5, 1, \ldots, 4\}$ m ($T_{60} = 0.3$ s, $N = 4$, non-coherent noise iSNR = 5 dB).

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.

9.6.3.2 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_{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. 9.7(a) the normalized global noise-reduction factor is shown for $0 \leq \alpha \leq 1$ and for input SNR (iSNR_{c}) of the coherent noise source that ranges from -5 dB to 30 dB. In Fig. 9.7(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.



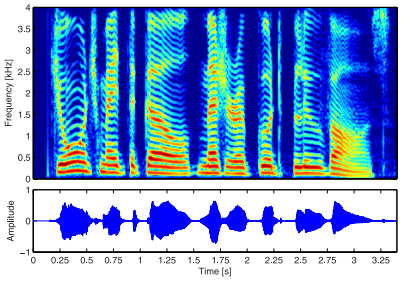
(a) (b)

Fig. 9.7 The normalized global noise-reduction factor obtained using a coherent plus non-coherent noise $\text{iSNR}_\text{c} = \{-5, \ldots, 30\}$ dB ($\text{iSNR}_\text{nc} = 20$ dB, $T_{60} = 0.3$ s, $N = 4$, $D = 2$ m).

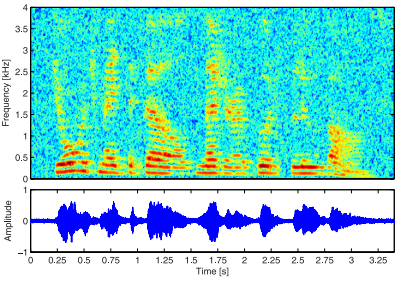
9.6.4 Example Using Speech Signals

Finally, we show an example obtained using a real speech sampled at 16 kHz. The speech sample was taken from the APLAWD speech database [28]. For this example non-coherent noise was used and the input SNR (iSNR_nc) was 10 dB. The distance between the source and the first microphone was set to $D = 3$ m and the reverberation time was set to $T_{60} = 0.35$ s. As shown in Subsection 9.6.1 there is a larger tradeoff between speech dereverberation and noise reduction when a limited amount of microphones is used. In order to emphasize the tradeoff we used four microphones. The AIRs were generated using the source-image method and are 1024 coefficients long. For this scenario long filters are required to estimate the direct response of the desired speech signal. The total length of the non-causal filter was 8192, of which 4096 coefficients correspond to the causal part of the filter. To avoid pre-echoes (i.e., echoes that arrive before the arrival of the direct sound), the non-causal part of the filter was properly truncated to a length of 1024 coefficients (128 ms). The filters and the second-order statistics of the signals are computed on a frame-by-frame basis in the discrete Fourier transform domain. The filter process is performed using the overlap-save technique [29].

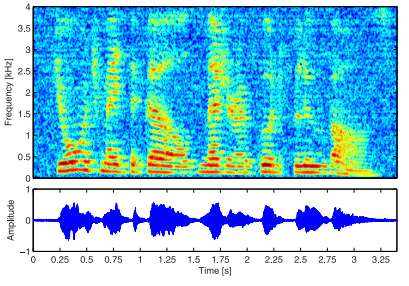
In Fig. 9.8(a) the spectrogram and waveform of the noisy and reverberant microphone signal $y_1(k)$ are depicted. In Fig. 9.8(b) the processed signal is shown with $\alpha = 1$, i.e., when desiring noise reduction only. Finally, in Fig. 9.8(c) the processed signal is shown with $\alpha = 0$, i.e. when desiring dereverberation and noise reduction. By comparing the spectrograms one can see that the processed signal shown in Fig. 9.8(c) contains less reverberation



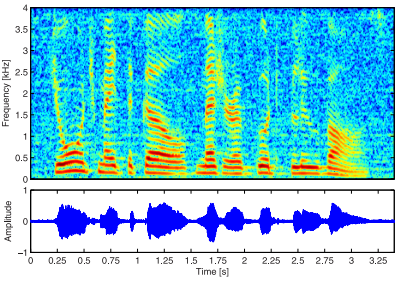
(a) The direct signal received by the first microphone.



(b) The microphone signal $y_1(k)$.



(c) The processed signal with $\alpha = 1$ (noise reduction only).



(d) The processed signal using $\alpha = 0$ (dereverberation and noise reduction).

Fig. 9.8 Spectrograms and waveforms of the unprocessed and processed signals ($\text{iSNR}_{\text{nc}} = 10$ dB, $T_{60} = 0.35$ s, $N = 4$, $D = 3$ m).

compared to the signals shown in Fig. 9.8(a) and Fig. 9.8(b). Specifically, the smearing in time is reduced and the harmonic structure of the speech are restored. In addition, we observe that there is a tradeoff between speech dereverberation and noise reduction as before. As expected, the processed signal in Fig. 9.8(c) contains more noise compared to the processed signal in Fig. 9.8(b).

9.7 Conclusions

In this chapter 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. 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

Proof (Property 9.1).

Before we proceed we define the magnitude squared coherence function (MSCF), which is the frequency-domain counterpart of the squared Pearson correlation coefficient (SPCC), which was used in [30] to analyze the noise reduction performance of the single-channel Wiener filter. Let $A(\omega)$ and $B(\omega)$ be the DTFTs of the two zero-mean real-valued random sequences a and b. Then the MSCF between $A(\omega)$ and $B(\omega)$ at frequency ω is defined as

$$\begin{aligned} \left|\rho\left[A(\omega),B(\omega)\right]\right|^2 &= \frac{\left|E\left[A(\omega)B^*(\omega)\right]\right|^2}{E\left[\left|A(\omega)\right|^2\right]E\left[\left|B(\omega)\right|^2\right]} \\ &= \frac{\left|\phi_{ab}(\omega)\right|^2}{\phi_{aa}(\omega)\phi_{bb}(\omega)}. \end{aligned} \tag{9.62}$$

It is clear that the MSCF always takes its values between zero and one.

Let us first evaluate the MSCF $\left|\rho\left[X_1(\omega),Y_1(\omega)\right]\right|^2$ [using (9.2) and (9.35)] and $\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^H(\omega)\mathbf{x}(\omega),\mathbf{h}_{\mathrm{MVDR}}^H(\omega)\mathbf{y}(\omega)\right]\right|^2$ [using (9.5) and (9.37)]:

$$\begin{aligned} \left|\rho\left[X_1(\omega),Y_1(\omega)\right]\right|^2 &= \frac{\left|G_1(\omega)\right|^2\phi_{ss}(\omega)}{\left|G_1(\omega)\right|^2\phi_{ss}(\omega)+\phi_{v_1v_1}(\omega)} \\ &= \frac{\mathrm{iSNR}\left[Q(\omega)\right]}{\frac{\left|Q(\omega)\right|^2}{\left|G_1(\omega)\right|^2}+\mathrm{iSNR}\left[Q(\omega)\right]}, \end{aligned} \tag{9.63}$$

$$\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{x}(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}=\frac{\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]}{1+\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]}. \quad (9.64)$$

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

$$\begin{aligned}\left|\rho\left[Y_1(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}&=\frac{\left|\mathbf{u}^{T}\boldsymbol{\Phi}_{yy}(\omega)\mathbf{h}_{\mathrm{MVDR}}(\omega)\right|^{2}}{\phi_{y_1y_1}(\omega)\cdot\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\boldsymbol{\Phi}_{yy}(\omega)\mathbf{h}_{\mathrm{MVDR}}(\omega)}\\&=\frac{\phi_{x_1x_1}(\omega)}{\phi_{y_1y_1}(\omega)}\cdot\frac{C(\omega)\phi_{x_1x_1}(\omega)}{\mathbf{u}^{T}\boldsymbol{\Phi}_{xx}(\omega)\mathbf{h}_{\mathrm{MVDR}}(\omega)}\\&=\frac{\left|\rho\left[X_1(\omega),Y_1(\omega)\right]\right|^{2}}{\left|\rho\left[X_1(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}}. \quad (9.65)\end{aligned}$$

From (9.65) and the fact that $|\rho[A(\omega),B(\omega)]|^2\leq 1$, we have

$$\begin{aligned}\left|\rho\left[X_1(\omega),Y_1(\omega)\right]\right|^{2}&=\left|\rho\left[Y_1(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}\times\\&\quad\left|\rho\left[X_1(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}\\&\leq\left|\rho\left[X_1(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}. \quad (9.66)\end{aligned}$$

In addition, it can be shown that

$$\begin{aligned}\left|\rho\left[X_1(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}&=\left|\rho\left[X_1(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{x}(\omega)\right]\right|^{2}\times\\&\quad\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{x}(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}\\&\leq\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{x}(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}. \quad (9.67)\end{aligned}$$

From (9.66) and (9.67), we know that

$$\left|\rho\left[X_1(\omega),Y_1(\omega)\right]\right|^{2}\leq\left|\rho\left[\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{x}(\omega),\mathbf{h}_{\mathrm{MVDR}}^{H}(\omega)\mathbf{y}(\omega)\right]\right|^{2}. \quad (9.68)$$

Hence, by substituting (9.63) and (9.64) in (9.68), we obtain

$$\frac{\mathrm{iSNR}\left[Q(\omega)\right]}{\frac{|Q(\omega)|^2}{|G_1(\omega)|^2}+\mathrm{iSNR}\left[Q(\omega)\right]}\leq\frac{\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]}{1+\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]}. \quad (9.69)$$

As a result

$$|Q(\omega)|^{2}\cdot\mathrm{oSNR}\left[\mathbf{h}_{\mathrm{MVDR}}(\omega)\right]\geq|G_1(\omega)|^{2}\cdot\mathrm{iSNR}\left[Q(\omega)\right],\ \forall\omega, \quad (9.70)$$

which is equal to (9.46).

□

References

1. J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*. Berlin, Germany: Springer-Verlag, 2008.
2. S. Gannot and I. Cohen, "Adaptive beamforming andpostfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2007, book chapter 48.
3. J. Chen, J. Benesty, Y. Huang, and S. Doclo, "New insights into the noise reduction Wiener filter," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 14, no. 4, pp. 1218–1234, July 2006.
4. S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multimicrophone speech enhancement," *IEEE Trans. Signal Process.*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.
5. A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction'," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.
6. J. Capon, "High resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, pp. 1408–1418, Aug. 1969.
7. S. Darlington, "Linear least-squares smoothing and prediction with applications," *Bell Syst. Tech. J.*, vol. 37, pp. 1121–94, 1952.
8. M. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 31, no. 6, pp. 1378–1393, Dec. 1983.
9. O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, Jan. 1972.
10. Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 6, pp. 1391–1400, Dec 1986.
11. 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.
12. 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, June 2002.
13. S. Affes and Y. Grenier, "A source subspace tracking array of microphones for double talk situations," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Munich, Germany, Apr. 1997, pp. 269–272.
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. 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.
16. Y. Huang, J. Benesty, and J. Chen, "Adaptive blind multichannel identification," in *Springer Handbook of Speech Processing*, J. Benesty, M. M. Sondhi, and Y. Huang, Eds. Springer-Verlag, 2007, book chapter 13.
17. S. Gannot and M. Moonen, "Subspace methods for multimicrophone speech dereverberation," *EURASIP Journal on Applied Signal Processing*, vol. 2003, no. 11, pp. 1074–1090, Oct. 2003.
18. E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, "New insights into the MVDR beamformer in room acoustics," *IEEE Trans. Audio, Speech, Language Process.*, 2010.
19. M. Brandstein and D. B. Ward, Eds., *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Germany: Springer-Verlag, 2001.
20. A. Oppenheim, A. Willsky, and H. Nawab, *Signals and Systems*. Upper Saddle River, NJ: Prentice Hall, 1996.

21. S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Frequency-domain criterion for the speech distortion weighted multichannel Wiener filter for robust noise reduction," *Speech Communication*, vol. 49, no. 7–8, pp. 636–656, Jul.–Aug. 2007.
22. A. Antoniou and W.-S. Lu, *Practical Optimization: Algorithms and Engineering Applications*. New York, USA: Springer, 2007.
23. H. Cox, R. Zeskind, and M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 35, no. 10, pp. 1365–1376, Oct. 1987.
24. J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, 1979.
25. E. E. Jan and J. Flanagan, "Sound capture from spatial volumes: Matched-filter processing of microphone arrays having randomly-distributed sensors," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Atlanta, Georgia, USA, May 1996, pp. 917–920.
26. G. Reuven, S. Gannot, and I. Cohen, "Performance analysis of the dual source transfer-function generalized sidelobe canceller," *Speech Communication*, vol. 49, pp. 602–622, Jul.–Aug. 2007.
27. P. M. Peterson, "Simulating the response of multiple microphones to a single acoustic source in a reverberant room," *Journal of the Acoustical Society of America*, vol. 80, no. 5, pp. 1527–1529, Nov. 1986.
28. G. Lindsey, A. Breen, and S. Nevard, "SPAR's archivable actual-word databases," University College London, Tech. Rep., Jun. 1987.
29. A. Oppenheim and R. W. Schafer, *Discrete-Time Signal Processing*, 2nd ed. Upper Saddle River, NJ: Prentice-Hall, Inc., 1999.
30. 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, pp. 757–765, May 2008.