CHAPTER 10

ACOUSTIC BEAMFORMING FOR HEARING AID APPLICATIONS

Simon Doclo, Sharon Gannot, Marc Moonen, Ann Spriet

10.1 INTRODUCTION

Noise reduction algorithms in hearing aids are crucial for hearing impaired persons to improve speech intelligibility in background noise (e.g. traffic, cocktail party situation). Many hearing aids currently have more than one microphone, enabling the use of multi-microphone speech enhancement algorithms [47]. In comparison with single-microphone algorithms, which can only use spectral and temporal information, multi-microphone algorithms can additionally exploit the spatial information of the sound sources. This generally results in a higher performance, especially when the speech and the noise sources are spatially separated.

Since many hearing impaired have a hearing loss at both ears, they are fitted with a hearing aid at each ear. In a so-called bilateral system, no cooperation between the hearing aids takes place. Current noise reduction algorithms in bilateral hearing aids are not designed to preserve the binaural localisation cues, i.e. the interaural time difference (ITD) and the interaural level difference (ILD) [99]. These binaural cues play an important role in sound localisation and speech segregation in noisy environments [6, 9, 12, 13, 48]. In order to achieve true binaural processing, both hearing aids need to cooperate with each other, e.g. through a wireless link, such that a binaural hearing aid can be considered a simple acoustic sensor network. The objective of a binaural signal enhancement algorithm then is not only to selectively extract the useful speech signal and suppress background noise, but also to preserve

Handbook on Array Processing and Sensor Networks. By K.J. Ray Liu, S. Haykin

Copyright © 2008 John Wiley & Sons, Inc.

the binaural cues of the sound sources, so as to preserve the auditory impression of the acoustic scene and exploit the binaural hearing advantage.

Section 10.2 gives an overview of several multi-microphone noise reduction techniques for hearing aids. In the remainder of the chapter, we focus on two specific techniques, i.e. adaptive beamforming and multi-channel Wiener filtering (MWF), and their use in monaural and binaural hearing aids. In Section 10.3 monaural beamforming techniques are discussed. We derive the standard generalised sidelobe canceler (GSC) and the transfer-function GSC from the minimum variance distortionless response beamformer, and we present the MWF and its relationship with the transfer-function GSC. In Section 10.4 MWF-based techniques for binaural hearing aids are discussed. We present extensions of the standard binaural MWF that specifically aim to preserve the binaural cues of the sound sources.

10.2 OVERVIEW OF NOISE REDUCTION TECHNIQUES

In the last decades, several multi-microphone speech enhancement algorithms for monaural and binaural hearing aids have been proposed, e.g. based on fixed and adaptive beamforming, multi-channel Wiener filtering (MWF), blind source separation (BSS) and computational auditory scene analysis (CASA). Each class of algorithms has its own advantages and limitations.

Fixed beamforming Fixed beamformers combine the microphone signals using a time-invariant filter-and-sum operation and are hence data-independent. The objective of a fixed beamformer is to obtain spatial focusing on the desired speech source, thereby reducing background noise not coming from the direction of the speech source. Different types of fixed beamformers exist, e.g. delay-and-sum beamforming, superdirective beamforming, differential microphone arrays and frequency-invariant beamforming [8, 30, 31, 35, 58, 59, 103]. A specific type of fixed beamformers, namely matched filtering, will be discussed in more detail in Section 10.3.2. For the design of fixed beamformers, the direction of the speech source and the complete microphone configuration need to be known. Hence, fixed beamformers have mainly been used for *monaural* hearing aids [60, 61, 90], although for *binaural* hearing aids fixed beamforming techniques have also been proposed that aim to combine spatial selectivity and noise reduction with the preservation of the binaural cues of the speech source [2, 26, 66, 72].

Adaptive beamforming In practice, since the background noise is unknown and can change both spectrally and spatially, information about the noise field has to be adaptively estimated. Adaptive beamformers combine the spatial focusing of fixed beamformers with adaptive noise suppression (e.g. adaptively steering a null in the direction of the dominant noise sources [67]). Hence, they generally exhibit a higher noise reduction performance than fixed beamformers.

In a minimum variance distortionless response (MVDR) beamformer [16, 39], the energy of the output signal is minimised under the constraint that signals arriving from the assumed direction of the desired speech source are processed without distortion. A widely studied adaptive implementation of this beamformer is the generalised sidelobe canceler (GSC) [46]. The standard GSC consists of a spatial pre-processor, i.e. a fixed beamformer and a blocking matrix, combined with a (multichannel) adaptive noise canceler (ANC). The fixed beamformer provides a spatial

focus on the speech source, creating a so-called speech reference; the blocking matrix steers nulls in the direction of the speech source, creating so-called noise references; and the ANC eliminates the noise components in the speech reference that are correlated with the noise references. Due to room reverberation, microphone mismatch and look direction error, speech components may however leak into the noise references of the standard GSC, giving rise to speech distortion and possibly signal cancelation. Several techniques have been proposed to limit the speech distortion resulting from this speech leakage, by reducing the speech leakage components in the noise references [17, 18, 40, 42, 51, 54, 76, 96] and by limiting the distorting effect of the remaining speech leakage components [23, 33, 52, 53, 56, 91, 94]. Several of these issues will be discussed in more detail in Section 10.3.

The GSC or one of its more robust variants is a widely used multi-microphone noise reduction technique for monaural hearing aids with an endfire microphone array configuration [45, 65, 67, 69, 102]. In an effort to combine adaptive noise reduction with binaural processing, adaptive beamforming techniques producing a binaural output signal have also been proposed. In [104] the frequency spectrum is divided into a low-pass and a high-pass portion, and the low-pass portion is passed through unaltered in order to preserve the ITD cues of the speech source, while adaptive noise reduction is performed only for the high-pass portion. Other algorithms restrict the preservation of the binaural cues to an angular region around the frontal direction, while reducing background noise for the other angles [75].

Multi-channel Wiener filtering (MWF) In [29] an MWF technique has been proposed that produces a minimum mean square error (MMSE) estimate of the desired speech component in one of the microphone signals, hence simultaneously performing noise reduction and limiting speech distortion. In addition, the MWF is able to take speech distortion into account in its optimisation criterion, resulting in the speech distortion weighted MWF (SDW-MWF). The SDW-MWF is uniquely based on estimates of the second-order statistics of the recorded speech signal and the noise signal, and hence requires a method that determines time-frequency regions where the desired source is dominant and time-frequency regions where the interference is dominant. The SDW-MWF has been successfully applied as a speech enhancement technique in monaural multi-microphone hearing aids [68, 93]. The monaural SDW-MWF and its relationship with adaptive beamforming will be discussed in more detail in Section 10.3.5.

Since the SDW-MWF produces an estimate of the speech component in the microphone signals and does not make any assumptions regarding the microphone configuration and the room impulse responses relating the speech source and the microphones, it is obviously well suited to combine noise reduction with binaural processing. It was shown in [28, 63] that the binaural MWF perfectly preserves the binaural cues of the speech component, but undesirably changes the noise cues to those of the speech component. Several extensions have been proposed to preserve the binaural cues of both the speech and the noise component, either by partial noise estimation [63] or by extending the MWF cost function with terms related to the ITD, ILD or interaural transfer function [27, 28, 97]. In addition, recently a distributed version of the binaural SDW-MWF has been presented [34]. The binaural MWF and its extensions will be discussed in more detail in Section 10.4.

Blind source separation (BSS) The signals received at the microphones can essentially be considered a mixture of all sound sources, filtered by the respective room

impulse responses between the sound source and the microphones. The goal of BSS is to recover all original signals. Many BSS algorithms exploit the independence and the non-Gaussianity of the sources, enabling the use of e.g. independent component analysis (ICA) techniques [4, 19]. While time-domain ICA-based techniques are well suited to solve the instantaneous mixing problem, they are not able to address the convolutive mixture problem encountered in typical reverberant environments. By considering the BSS problem in the frequency-domain, the convolutive mixing problem can be transformed into an instantaneous mixing problem for each frequency bin. An inherent permutation and amplitude/phase scaling problem however occurs in frequency-domain BSS approaches, for which several solutions have been proposed [15, 70, 81]. Nevertheless, due to its computational complexity the use of BSS techniques in hearing aids has found only limited practical interest.

Computational auditory scene analysis Algorithms based on computational auditory scene analysis (CASA) aim to perform sound segregation by modeling the human auditory perceptual processing [12, 14, 36, 86]. A typical CASA model involves two stages. In the first stage, a time-frequency representation of the incoming mixture of signals is generated, e.g. using short-time Fourier transform (STFT) processing or by incorporating a more advanced cochlear model. In the second stage, the resulting time-frequency elements are grouped into separate perceptual streams based on distinctive perceptual cues. As summarised in [12], some cues characterise the monaural acoustic properties of the sources, such as common pitch, amplitude modulation, onset [55, 64]. In addition, for a binaural system sound sources can also be distinguished based on their spatial direction information, e.g. ITD, ILD and interaural envelope difference [73, 84, 105]. A gain factor is applied to each time-frequency element, such that regions dominated by the desired sound stream receive a high gain and regions dominated by other streams receive a low gain.

10.3 MONAURAL BEAMFORMING

10.3.1 Problem Formulation

Consider an array of M microphones in a noisy and reverberant environment. The received signals consist of two components: a desired speech signal and an undesired noise signal. The observed signal $x_m(n)$ at the mth microphone is hence given by

$$x_m(n) = \sum_{i=0}^{L_h} h_{m,i}(n)s(n-i) + v_m(n) = s_m(n) + v_m(n)\,, \quad m = 1, \ldots, M\,. \quad (10.1)$$

where $s_m(n)$ and $v_m(n)$ represent the speech and the noise component at the mth microphone and n is the time index. It is assumed that the noise statistics are slowly time-varying, while the speech signal is non-stationary. The noise signal may comprise coherent (directional) as well as non-coherent (diffuse and uncorrelated) noise components. The room impulse responses (RIRs) $h_{m,i}(n)$, relating the speech source with the mth microphone, are in general time-varying finite impulse response (FIR) filters of order L_h, where $i = 0, \ldots, L_h$ is the filter tap index. We will assume in the remainder of the chapter that the RIRs are slowly time-varying, i.e. $h_{m,i}(n) \approx h_{m,i}$.

The observed signals are analysed using the STFT, yielding

$$x_m(k,\ell) \approx h_m(k)s(k,\ell) + v_m(k,\ell) \;, \tag{10.2}$$

where $x_m(k,\ell), s(k,\ell)$ and $v_m(k,\ell)$ are the STFT of the respective time-domain signals[1], ℓ represents the frame index and $k = 0, 1, \ldots, K-1$ represents the frequency bin index[2]. $h_m(k)$, the Fourier transform of the RIR $h_{m,i}$, is denoted acoustical transfer function (ATF). It is clearly seen that in the STFT representation each frequency bin can be treated separately. Hence, for the compactness of the exposition, the dependency on the frequency bin index k will be omitted from this point on. In addition, the dependency on the frame index ℓ will be omitted, except for the adaptive algorithms and the relative transfer function estimation in Section 10.3.4.

The equation set (10.2) can be stated in a vector form as

$$\mathbf{x} = \mathbf{h}s + \mathbf{v} \;, \tag{10.3}$$

with

$$\mathbf{x} = \begin{bmatrix} x_1 & x_2 & \cdots & x_M \end{bmatrix}^T \;, \tag{10.4}$$

$$\mathbf{h} = \begin{bmatrix} h_1 & h_2 & \cdots & h_M \end{bmatrix}^T \;, \tag{10.5}$$

$$\mathbf{v} = \begin{bmatrix} v_1 & v_2 & \cdots & v_M \end{bmatrix}^T \;. \tag{10.6}$$

A beamformer, depicted in Figure 10.1, is a system processing each of the microphone signals x_m by the filters w_m^* and summing the outputs. Define

$$\mathbf{w}^H = \begin{bmatrix} w_1^* & w_2^* & \cdots & w_M^* \end{bmatrix} \;, \tag{10.7}$$

where ^{H} denotes conjugate transpose. The beamformer output in the STFT domain is then given by

$$y = \mathbf{w}^H\mathbf{x} = \mathbf{w}^H\mathbf{h}s + \mathbf{w}^H\mathbf{v} \stackrel{\triangle}{=} y_s + y_v \;, \tag{10.8}$$

where y_s and y_v represent the speech and the noise component at the output of the beamformer. Finally, the output signal in the time-domain $y(n)$ is reconstructed using the inverse short-time Fourier transform (iSTFT).

The signal-to-noise ratio (SNR) of the beamformer output is defined as

$$\mathrm{SNR}^{out} \stackrel{\triangle}{=} \frac{E\{|y_s|^2\}}{E\{|y_v|^2\}} = \frac{|\mathbf{w}^H\mathbf{h}|^2\phi_{ss}}{\mathbf{w}^H\mathbf{\Phi}_{\mathbf{vv}}\mathbf{w}} \;, \tag{10.9}$$

with $\phi_{ss} = E\{|s|^2\}$ the power spectral density (PSD) of the speech signal and $\mathbf{\Phi}_{\mathbf{vv}} \stackrel{\triangle}{=} E\{\mathbf{v}\mathbf{v}^H\}$ the correlation matrix of the noise components measured at the microphone signals.

[1]We use lowercase symbols to denote both time-domain and frequency-domain signals. The signals are distinguished by their functional dependency: time-domain signals depend on the discrete time index n, whereas frequency-domain signals depend on the frame index ℓ and the frequency bin index k (which are frequently omitted).

[2]Equality in (10.2) is only valid for segments that are significantly longer than the RIR length, i.e. the number of frequency bins K is assumed to be larger than the STFT segment length. Since RIRs tend to be very long, the conditions allowing for this representation to hold cannot be exactly met. We assume, however, that the STFT relation is a reasonable approximation.

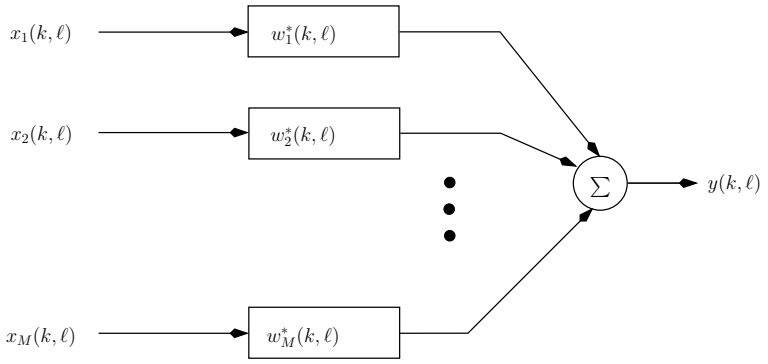


Figure 10.1 Filter-and-sum beamformer

10.3.2 Data-Independent Beamformer

The simplest beamformer structure is the *delay-and-sum beamformer*, which first compensates for the relative delay between the microphone signals and then sums the steered signals to form a single output. This beamformer, which is still widely used, can be very effective in mitigating non-coherent, i.e. spatially white, noise sources, provided that the number of microphones is relatively high. However, if the noise source is coherent, the noise reduction performance is strongly dependent on the direction of arrival of the noise source. Consequently, the performance of the delay-and-sum beamformer in reverberant environments is often insufficient. The delay-and-sum concept has been extended by introducing the *matched filter beamformer* (MBF) [57, 58, 82]. This structure, designed for multi-path environments such as reverberant enclosures, replaces the simple delay-compensator by a matched filter, i.e.

$$\mathbf{w}_0 = \frac{\mathbf{h}}{\|\mathbf{h}\|^2} \,. \tag{10.10}$$

Since the array beam pattern of the delay-and-sum beamformer and the MBF do not depend on the measured signals, they are often referred to as *data-independent* beamformers. It should be noted that the application of data-independent design methods is very limited in dynamic acoustical environments.

10.3.3 Minimum Variance Distortionless Response (MVDR) Beamformer

The MVDR beamformer, proposed by Capon [16], constrains the response of the beamformer in the desired look direction while minimising the response in all other directions. The MVDR concept was extended to satisfy a set of linear constraints, resulting in the linearly constrained minimum variance (LCMV) beamformer (for details we refer to [23, 101]). Frost [39] proposed an adaptive form of Capon's beamformer, which relies on the assumption that the RIRs, relating the desired source and the microphones, can be uniquely represented by gain and delay values.

In this section, we derive the MVDR beamformer for the general reverberant case, and show that under the distortionless response constraint it is equivalent to the maximum signal-to-noise ratio (MSNR) beamformer.

Derivation Using (10.8), the output PSD of the beamformer is given by

$$E\{|y|^2\} = E\{\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}\} = \mathbf{w}^H\mathbf{\Phi}_{\mathbf{xx}}\mathbf{w} \,, \tag{10.11}$$

where $\mathbf{\Phi}_{\mathbf{xx}} \stackrel{\triangle}{=} E\{\mathbf{x}\mathbf{x}^H\}$ is the correlation matrix of the received microphone signals. We want to minimise the output PSD subject to the constraint that y_s, the speech component at the beamformer output, is equal to the speech source signal s, i.e.

$$y_s = \mathbf{w}^H\mathbf{h}s \stackrel{\text{constrain}}{=} s \,. \tag{10.12}$$

Hence, the constrained minimisation problem can be stated as

$$\min_{\mathbf{w}} \mathbf{w}^H\mathbf{\Phi}_{\mathbf{xx}}\mathbf{w} \,, \text{ subject to } \mathbf{w}^H\mathbf{h} = 1 \,. \tag{10.13}$$

To solve (10.13), we first define the complex Lagrangian, i.e.

$$\mathcal{L} = \mathbf{w}^H\mathbf{\Phi}_{\mathbf{xx}}\mathbf{w} + \lambda\left[\mathbf{w}^H\mathbf{h} - 1\right] + \lambda^*\left[\mathbf{h}^H\mathbf{w} - 1\right] \,, \tag{10.14}$$

where λ is a Lagrange multiplier. Setting the derivative with respect to $\mathbf{w}^*$ to 0 yields

$$\nabla_{\mathbf{w}^*}\mathcal{L} = \mathbf{\Phi}_{\mathbf{xx}}\mathbf{w} + \lambda\mathbf{h} = 0 \,. \tag{10.15}$$

Imposing the constraint in (10.13), we obtain the MVDR filter

$$\mathbf{w}_{\text{MVDR}} = \frac{\mathbf{\Phi}_{\mathbf{xx}}^{-1}\mathbf{h}}{\mathbf{h}^H\mathbf{\Phi}_{\mathbf{xx}}^{-1}\mathbf{h}} \,. \tag{10.16}$$

Equivalence between MVDR and MSNR beamformer It is interesting to show the equivalence between the MVDR solution (10.16) and the MSNR beamformer [23], which is obtained by maximising the output SNR in (10.9). The well-known solution to this maximisation is the (coloured noise) matched filter, i.e. $\mathbf{w} \propto \mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}$. If the array response is constrained to satisfy $\mathbf{w}^H\mathbf{h} = 1$, i.e. signals impinging the array from the look direction suffer no distortion, we obtain

$$\mathbf{w}_{\text{MSNR}} = \frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}} \,. \tag{10.17}$$

Using (10.3), it is evident that

$$\mathbf{\Phi}_{\mathbf{xx}} = \phi_{ss}\mathbf{h}\mathbf{h}^H + \mathbf{\Phi}_{\mathbf{vv}} \,. \tag{10.18}$$

To show the equivalence of the MVDR beamformer (10.16) and the MSNR beamformer (10.17), we apply the matrix inversion lemma for inverting $\mathbf{\Phi}_{\mathbf{xx}}$, i.e.

$$\mathbf{\Phi}_{\mathbf{xx}}^{-1} = \left(\phi_{ss}\mathbf{h}\mathbf{h}^H + \mathbf{\Phi}_{\mathbf{vv}}\right)^{-1} = \mathbf{\Phi}_{\mathbf{vv}}^{-1} - \frac{\phi_{ss}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}}{1+\rho} \,, \tag{10.19}$$

with

$$\rho = \phi_{ss}\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h} \,, \tag{10.20}$$

such that

$$\left(\phi_{ss}\mathbf{h}\mathbf{h}^H + \mathbf{\Phi}_{\mathbf{vv}}\right)^{-1}\mathbf{h} = \frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{1+\rho} \,, \tag{10.21}$$

$$\mathbf{h}^H\left(\phi_{ss}\mathbf{h}\mathbf{h}^H + \mathbf{\Phi}_{\mathbf{vv}}\right)^{-1}\mathbf{h} = \frac{\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{1+\rho} \,. \tag{10.22}$$

Dividing (10.21) by (10.22), we obtain

$$\mathbf{w}_{\mathrm{MVDR}} = \frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}} \equiv \mathbf{w}_{\mathrm{MSNR}} \, . \tag{10.23}$$

Hence, using (10.9), also the output SNR of both beamformers is identical, i.e.

$$\mathrm{SNR}_{\mathrm{MVDR}}^{out} = \mathrm{SNR}_{\mathrm{MSNR}}^{out} = \phi_{ss}\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h} = \rho \, , \tag{10.24}$$

which obviously depends on the input SNR and on the spatial separation between the speech source and the noise sources. While both MSNR and MVDR beamformers are shown to be equivalent, provided that the ATFs $\mathbf{h}$ are known, their behaviour in case of unknown ATFs is different and has been analysed in [22].

10.3.4 Frequency-Domain Generalised Sidelobe Canceler

Griffiths and Jim [46] proposed a method to split the constrained minimisation (10.13) in the MVDR formulation into two orthogonal operations. Their beamformer structure is referred to as the generalised sidelobe canceler (GSC) and consists of two branches. The first branch, satisfying the look direction constraint, is a delay-and-sum beamformer. The second branch, which corresponds to an unconstrained minimisation of the beamformer output power, mitigates all signals impinging the array from directions other than the look direction. Breed and Strauss [11] proved that an equivalent GSC structure also exists for the more general LCMV beamformer.

In this section, we reformulate the derivation of Griffiths and Jim in the frequency-domain for arbitrary ATFs relating the speech source and the microphones. A more detailed description of the resulting transfer-function GSC (TF-GSC) can be found in [40]. This GSC structure is also used in [83] to deal with the more involved problem, in which an additional non-stationary interference is received by the array. By imposing an additional linear constraint on the beamformer output, the resulting dual source TF-GSC enhances the interference cancelation.

Derivation Consider the $M \times M$-dimensional full-rank matrix $\mathbf{U} = \left[\begin{array}{cc} \mathbf{h} & \mathbf{B} \end{array} \right]$, where the columns of the matrix $\mathbf{B}$ span the null space of $\mathbf{h}$, i.e.

$$\mathbf{h}^H\mathbf{B} = 0 \, , \quad \mathrm{rank}\,\{\mathbf{B}\} = M - 1 \, . \tag{10.25}$$

The matrix $\mathbf{B}$ is usually denoted blocking matrix (BM). Define

$$\left[\begin{array}{c} g_1 \\ \mathbf{g} \end{array} \right] = -\mathbf{U}^{-1}\mathbf{w} \, , \tag{10.26}$$

such that the filter $\mathbf{w}$ can be uniquely split as

$$\mathbf{w} = -g_1\mathbf{h} - \mathbf{B}\mathbf{g} \, , \tag{10.27}$$

with $\mathbf{g}$ an $(M-1)$-dimensional filter. By imposing the constraint $\mathbf{w}^H\mathbf{h} = 1$, we obtain $g_1 = -1/\|\mathbf{h}\|^2$, such that

$$\mathbf{w} = \mathbf{w}_0 - \mathbf{B}\mathbf{g} \, , \tag{10.28}$$

with $\mathbf{w}_0$ the matched beamformer in (10.10), which depends on the ATFs. The criterion in (10.13) can now be solved by a simple unconstrained minimisation of the cost function with respect to $\mathbf{g}$ rather than $\mathbf{w}$.

The output of the constrained beamformer is given by

$$y = \mathbf{w}_0^H \mathbf{x} - \mathbf{g}^H \mathbf{B}^H \mathbf{x} \triangleq y_{\text{MBF}} - y_{\text{ANC}} \,. \tag{10.29}$$

In addition, we define the $M-1$ output signals of the blocking matrix

$$\mathbf{u} \triangleq \mathbf{B}^H \mathbf{x} = \mathbf{B}^H (\mathbf{h}s + \mathbf{v}) = \mathbf{B}^H \mathbf{v} \,, \tag{10.30}$$

which are referred to as *noise reference signals*, as they only contain noise components.

The beamformer output y is the difference of two terms, both operating on the input signal $\mathbf{x}$. The first term, y_{MBF}, is the output of the matched beamformer. The MBF coherently sums the desired speech components, while in general destructively summing the noise components. Hence, it is expected that the SNR at the MBF output is higher than the input SNR (for a comprehensive discussion of this issue, we refer to [41]). The second term, y_{ANC}, is obtained by filtering the noise reference signals $\mathbf{u}$ with the adaptive noise cancelation (ANC) filters $\mathbf{g}$.

The residual noise term in y_{MBF} can be reduced by properly adjusting the filters $\mathbf{g}$, by solving the unconstrained minimisation problem

$$\min_{\mathbf{g}} \left\{ [\mathbf{w}_0 - \mathbf{B}\mathbf{g}]^H \, \mathbf{\Phi}_{\mathbf{vv}} \, [\mathbf{w}_0 - \mathbf{B}\mathbf{g}] \right\} \,. \tag{10.31}$$

The solution is given by (see also [7, 78])

$$\mathbf{g} = \left[\mathbf{B}^H \mathbf{\Phi}_{\mathbf{vv}} \mathbf{B} \right]^{-1} \mathbf{B}^H \mathbf{\Phi}_{\mathbf{vv}} \mathbf{w}_0 \,. \tag{10.32}$$

In practice, this unconstrained minimisation problem can be adaptively solved using e.g. the normalised least mean squares (NLMS) algorithm [49], i.e.

$$\mathbf{g}(k, \ell+1) = \begin{cases} \mathbf{g}(k,\ell) + \frac{\lambda}{p_{\text{est}}(k,\ell)} \mathbf{u}(k,\ell) y^*(k,\ell) & \text{inactive speech periods,} \\ \mathbf{g}(k,\ell) & \text{otherwise,} \end{cases} \tag{10.33}$$

where

$$p_{\text{est}}(k,\ell) = \alpha p_{\text{est}}(k,\ell-1) + (1-\alpha) \|\mathbf{u}(k,\ell)\|^2 \tag{10.34}$$

represents the PSD estimate of the noise reference signals, λ is a step-size that regulates the convergence rate, and α is a smoothing parameter in the PSD estimation process. To allow for the use of the STFT, we further constrain the ANC filters $\mathbf{g}$ to have a (time-varying) FIR structure $g_{m,i}(n) = 0, \; i \notin \{-L_{g,\text{left}}, \ldots, L_{g,\text{right}}\}$. Note that the impulse responses are typically taken to be non-causal, to allow for relative delays between the MBF and the ANC branches. The various filtering (multiplications in the STFT domain) operations involved in the algorithm can be realised using the overlap-save method [24, 88].

Suboptimal GSC The frequency-domain GSC, derived in the previous section, is solely determined by the ATFs $\mathbf{h}$. Due to the large order of the respective RIRs, estimating these ATFs generally is a cumbersome task. Alternatively, the constraint

for a distortionless response can be replaced by constraining the output speech component to be equal to the speech component at an arbitrarily chosen microphone r. Hence, the constraint in (10.12) becomes

$$y_s = \mathbf{w}^H \mathbf{h} s \overset{\text{constrain}}{=} s_r = s h_r \ , \tag{10.35}$$

which is equivalent to $\mathbf{w}^H \tilde{\mathbf{h}} = 1$, where the relative transfer function (RTF) $\tilde{\mathbf{h}}$ is defined as

$$\tilde{\mathbf{h}} \overset{\triangle}{=} \frac{\mathbf{h}}{h_r} = \left[\begin{array}{ccccc} \frac{h_1}{h_r} & \frac{h_2}{h_r} & \cdots & 1 & \cdots \frac{h_M}{h_r} \end{array} \right]^T \ . \tag{10.36}$$

Note that all ATFs, and in particular h_r, may have zeros outside the unit circle, as they are not necessarily minimum-phase systems. Hence, to ensure the stability of the RTFs, we allow for non-causal systems and model the impulse responses of the RTFs $\tilde{\mathbf{h}}$ as non-causal FIR filters. Explicitly, for $m = 1, \ldots, M$; $m \neq r$ $\tilde{h}_{m,i}(n) = 0$, $i \notin \{-L_{\tilde{h},\text{left}}, \ldots, L_{\tilde{h},\text{right}}\}$, and for $m = r$, $\tilde{h}_m(n) = \delta(n)$. It was shown experimentally that the RTFs are usually much shorter than the corresponding ATFs [40], justifying the FIR assumption.

By substituting the ATFs $\mathbf{h}$ with the RTFs $\tilde{\mathbf{h}}$, the modified MVDR and MSNR beamformers are obtained, i.e.

$$\tilde{\mathbf{w}}_{\text{MVDR}} = h_r^* \mathbf{w}_{\text{MVDR}} = \frac{\mathbf{\Phi}_{\mathbf{xx}}^{-1} \tilde{\mathbf{h}}}{\tilde{\mathbf{h}}^H \mathbf{\Phi}_{\mathbf{xx}}^{-1} \tilde{\mathbf{h}}} = \frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1} \tilde{\mathbf{h}}}{\tilde{\mathbf{h}}^H \mathbf{\Phi}_{\mathbf{vv}}^{-1} \tilde{\mathbf{h}}} = \tilde{\mathbf{w}}_{\text{MSNR}} \ . \tag{10.37}$$

Similarly, a (sub-optimal) modified frequency-domain GSC, referred to as TF-GSC [40], can be constructed, in which the distortionless response property is sacrificed for the simplicity of the structure. In particular, the MBF is given by

$$\tilde{\mathbf{w}}_0 = \frac{\tilde{\mathbf{h}}}{\|\tilde{\mathbf{h}}\|^2} \ , \tag{10.38}$$

such that, using (10.3) and (10.29), the MBF output is equal to

$$y_{\text{MBF}} = h_r s + \frac{\tilde{\mathbf{h}}^H}{\|\tilde{\mathbf{h}}\|^2} \mathbf{v} \ . \tag{10.39}$$

Hence, when using $\tilde{\mathbf{w}}_0$ rather than $\mathbf{w}_0$, the speech component in y_{MBF} becomes $s_r = h_r s$, as expected. Note that all microphone signals are still added coherently.

The RTFs can also be used for constructing the blocking matrix. To show this, define the noise reference signals u_m, $m = 1, 2, \ldots, M$; $m \neq r$, as

$$u_m = x_m - \tilde{h}_m x_r = h_m s + v_m - \frac{h_m}{h_r} (h_r s + v_r) = v_m - \tilde{h}_m v_r \ , \tag{10.40}$$

which contain only noise components, provided that $\tilde{\mathbf{h}}$ correctly models the RTFs. The elements of the blocking matrix can be readily extracted from $\mathbf{u} = \mathbf{B}^H \mathbf{x}$, i.e.

$$\mathbf{B} = \left[\begin{array}{cccc} -\tilde{h}_2^* & -\tilde{h}_3^* & \ldots & -\tilde{h}_M^* \\ 1 & 0 & \ldots & 0 \\ 0 & 1 & \ldots & 0 \\ & \ldots & \ddots & \\ 0 & 0 & \ldots & 1 \end{array} \right] \ . \tag{10.41}$$

The consequences of using the RTFs rather than the ATFs in the GSC structure are more evident now. On the one hand, since the output speech component is equal to the speech component at the reference microphone, the proposed sub-optimal TF-GSC [40] does not aim at speech dereverberation. On the other hand, as no speech leaks to the noise reference signals, no speech self-cancelation occurs and no annoying speech distortion artifacts are expected. This is not the case when using the delay-only model for the RIRs as in the regular GSC [46], typically leading to severe distortion effects.

Relative Transfer Function Estimation This section discusses how the RTFs $\tilde{\mathbf{h}}$ can be estimated in practice using the system identification technique proposed in [87] and later used in the context of microphone arrays [40]. This method relies on the assumptions that the background noise is stationary, that the desired signal $s(n)$ is non-stationary, and that the support of the relative impulse responses relating the microphones is finite and slowly time-varying.

Rearranging the terms in (10.40), we have

$$x_m = \tilde{h}_m x_r + u_m \; . \tag{10.42}$$

Collecting L_i STFT frames $\ell_i, \ldots, \ell_i + L_i - 1$ for $i = 1, \ldots, I$, we obtain I different estimates of the time-varying cross-PSDs between the microphone signals (for each frequency bin). Using (10.42), we have

$$\phi^{(i)}_{x_m x_r} = \tilde{h}_m \phi^{(i)}_{x_r x_r} + \phi_{u_m x_r} \; , \qquad i = 1, \ldots, I \; , \tag{10.43}$$

where $\phi^{(i)}_{x_m x_r}$ is the cross-PSD between x_m and x_r and $\phi^{(i)}_{x_r x_r}$ is the PSD of x_r during the ith segment. The cross-PSD $\phi_{u_m x_r}$ between u_m and x_r is assumed to be independent of the segment index i due to the noise stationarity. Let $\hat{\phi}^{(i)}_{x_r x_r}$, $\hat{\phi}^{(i)}_{x_m x_r}$ and $\hat{\phi}^{(i)}_{u_m x_r}$ be estimates of $\phi^{(i)}_{x_r x_r}$, $\phi^{(i)}_{x_m x_r}$ and $\phi_{u_m x_r}$, obtained by replacing expectations with averages. Note that (10.43) still holds for the estimated values. Let $\varepsilon^{(i)}_m = \hat{\phi}^{(i)}_{u_m x_r} - \phi_{u_m x_r}$ denote the estimation error of the cross-PSD between u_m and x_r during the ith segment. Hence,

$$\hat{\phi}^{(i)}_{x_m x_r} = \tilde{h}_m \hat{\phi}^{(i)}_{x_r x_r} + \phi_{u_m x_r} + \varepsilon^{(i)}_m \; , \qquad i = 1, \ldots, I \; . \tag{10.44}$$

If the noise reference signals u_m were uncorrelated with x_r, then the standard system identification estimate

$$\hat{\tilde{h}}_m = \hat{\phi}_{x_m x_r} / \hat{\phi}_{x_r x_r} \tag{10.45}$$

could be used to obtain an unbiased estimate of $\tilde{h}_m$. Unfortunately, by (10.40) u_m and x_r are in general correlated. Hence, it is proposed in [87] to obtain an unbiased estimate of $\tilde{h}_m$ by applying a weighted least squares procedure to the set of overdetermined equations (10.44), where a separate set of equations is used for each $m = 1, \ldots, M$; $m \neq r$.

Shalvi and Weinstein used a diagonal weight matrix, in which the weights are proportional to L_i, the number of STFT frames used for estimating the PSDs, yielding

$$\hat{\tilde{h}}_m = \frac{\left\langle \hat{\phi}_{x_m x_r} \, \hat{\phi}_{x_r x_r} \right\rangle - \left\langle \hat{\phi}_{x_m x_r} \right\rangle \left\langle \hat{\phi}_{x_r x_r} \right\rangle}{\left\langle \hat{\phi}^2_{x_r x_r} \right\rangle - \left\langle \hat{\phi}_{x_r x_r} \right\rangle^2} \; , \tag{10.46}$$

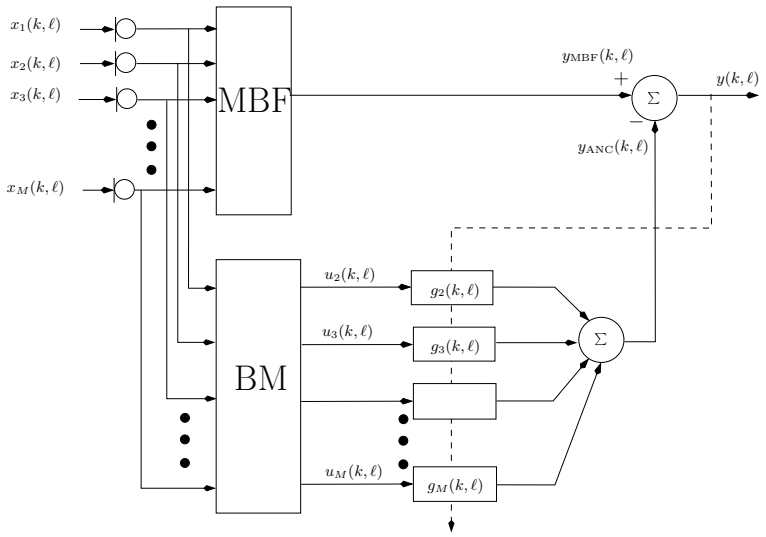


Figure 10.2 Linearly constrained adaptive beamformer.

with the averaging operation defined as

$$\langle \varphi \rangle \triangleq \frac{\sum_{i=1}^{I} L_i \varphi^{(i)}}{\sum_{i=1}^{I} L_i} \,. \tag{10.47}$$

Special attention should be given to choosing L_i. On the one hand, it should be longer than the correlation length of $x_m(n)$, which must be longer than the length of the corresponding RIR. On the other hand, it should be short enough for the filter time-invariance and the noise stationarity assumptions to hold.

Summary In this section, we surveyed the TF-GSC [40], which extends the original GSC [46] for dealing with arbitrary ATFs. Figure 10.2 depicts a block diagram of the TF-GSC, which comprises three parts: a matched beamformer (MBF) $\tilde{\mathbf{w}}_0$, which aligns the desired speech component at each microphone; a blocking matrix (BM) $\mathbf{B}$, which blocks the desired speech component resulting in the noise reference signals $\mathbf{u}$; and a multichannel ANC $\mathbf{g}$, which eliminates the residual noise component at the MBF output. The steps involved in the computation are summarised in Algorithm 1, where the operation $\overset{\text{FIR}}{\longleftarrow}$ imposes an FIR structure on the adaptive filters to allow for the application of the overlap-save procedure [88].

10.3.5 Multi-channel Wiener Filter (MWF)

MWF techniques for noise reduction have been proposed that produce a minimum mean square error (MMSE) estimate of either the speech source signal [71, 89], the (reverberant) speech component in one of the microphone signals [29, 77, 79, 85, 92] or a reference speech signal [1, 77, 79]. In this section, we will focus on MWF techniques that produce an estimate of the speech component in one of the microphone signals [29, 85]. Similarly to the TF-GSC, these MWF techniques

Algorithm Summary of the TF-GSC algorithm.

1) Matched beamformer:
$$y_{\mathrm{MBF}}(k,\ell) = \tilde{\mathbf{w}}_0^H(k,\ell)\mathbf{x}(k,\ell)$$
$\tilde{\mathbf{w}}_0$ is defined in (10.38) and the RTFs are estimated using (10.46).
2) Noise reference signals:
$$\mathbf{u}(k,\ell) = \mathbf{B}^H(k,\ell)\mathbf{x}(k,\ell)$$
$\mathbf{B}$ is constructed using (10.41) and the RTFs are estimated using (10.46).
3) Output signal:
$$y(k,\ell) = y_{\mathrm{MBF}}(k,\ell) - \mathbf{g}^H(k,\ell)\mathbf{u}(k,\ell)$$
4) Filters update:
$\tilde{\mathbf{g}}(k,\ell+1) = \mathbf{g}(k,\ell) + \frac{\lambda}{p_{\mathrm{est}}(k,\ell)}\mathbf{u}(k,\ell)y^*(k,\ell)$
$\mathbf{g}(k,\ell+1) \overset{\mathrm{FIR}}{\longleftarrow} \tilde{\mathbf{g}}(k,\ell+1)$
where $p_{\mathrm{est}}(k,\ell) = \alpha p_{\mathrm{est}}(k,\ell-1) + (1-\alpha)\|\mathbf{u}(k,\ell)\|^2$
5) keep only non-aliased samples [88].

neither require a priori information about the microphone configuration nor the position of the desired speech source, making it an appealing approach from a robustness point of view [29, 93]. These MWF techniques are uniquely based on estimates of the second-order statistics of the speech and the noise signal.

Concept: MMSE-criterium The M-dimensional MWF $\mathbf{w}_{\mathrm{MWF}}$ minimises the mean square error between the (unknown) speech component $s_r = h_r s$ in the rth microphone signal and the beamformer output, i.e.

$$\mathbf{w}_{\mathrm{MWF}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, E\{|s_r - y|^2\} = \underset{\mathbf{w}}{\operatorname{argmin}}\, E\{|s_r - \mathbf{w}^H\mathbf{x}|^2\}\,. \tag{10.48}$$

This is illustrated in Figure 10.3. Using (10.3) and assuming the speech and the noise components to be uncorrelated, the solution of (10.48) is equal to[3]

$$\mathbf{w}_{\mathrm{MWF}} = \mathbf{\Phi}_{\mathbf{xx}}^{-1}E\{\mathbf{x}s_r^*\} = \left(\phi_{ss}\mathbf{h}\mathbf{h}^H + \mathbf{\Phi}_{\mathbf{vv}}\right)^{-1}\phi_{ss}\mathbf{h}h_r^*\,. \tag{10.49}$$

Speech distortion weighted MWF The MMSE criterion (10.48) can be easily generalised to allow for a trade-off between noise reduction and speech distortion [33, 91]. We refer to this generalisation as speech distortion weighted MWF (SDW-MWF). The residual energy $E\{|s_r - \mathbf{w}^H\mathbf{x}|^2\}$ of the MWF can be decomposed into

$$E\{|s_r - \mathbf{w}^H\mathbf{h}s|^2\} + E\{|\mathbf{w}^H\mathbf{v}|^2\}\,, \tag{10.50}$$

where $E\{|s_r - \mathbf{w}^H\mathbf{h}s|^2\}$ represents the speech distortion energy and $E\{|\mathbf{w}^H\mathbf{v}|^2\}$ represents the residual noise energy. Hence, the design criterion (10.48) of the MWF can be easily generalised by incorporating a weighting factor μ [29], i.e.

$$\mathbf{w}_{\mathrm{SDW}} = \underset{\mathbf{w}}{\operatorname{argmin}}\, E\{|s_r - \mathbf{w}^H\mathbf{h}s|^2\} + \mu E\{|\mathbf{w}^H\mathbf{v}|^2\}\,. \tag{10.51}$$

[3]In (10.49), a rank-one speech correlation matrix is assumed. However, the MWF can also be defined more generally for a full-rank speech correlation matrix [29].

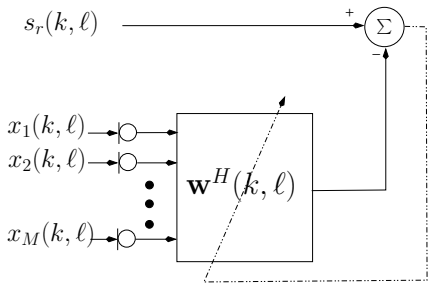


Figure 10.3 Multi-channel Wiener Filter (MWF).

The solution of (10.51) is given by

$$\mathbf{w}_{\mathrm{SDW}} = \left(\phi_{ss}\mathbf{h}\mathbf{h}^H + \mu\mathbf{\Phi_{vv}}\right)^{-1}\phi_{ss}\mathbf{h}h_r^* \,. \tag{10.52}$$

The factor $\mu \in [0, \infty]$ allows to trade off speech distortion versus noise reduction, hence, the name SDW-MWF: the smaller μ, the smaller the resulting speech distortion. If $\mu = 1$, the MMSE criterion (10.48) is obtained. If $\mu > 1$, the residual noise level will be reduced at the expense of increased speech distortion. By setting μ to ∞, all emphasis is put on noise reduction and speech distortion is completely ignored, resulting in $\mathbf{w} = 0$. On the other hand, setting μ to 0 results in $\mathbf{w} = \mathbf{e}_r$, with the M-dimensional vector $\mathbf{e}_r$ equal to the rth canonical vector, i.e.

$$\mathbf{e}_r = \left[\begin{array}{ccccccccc} 0 & \cdots & 0 & 1 & 0 & \cdots & 0 \end{array}\right]^T \,, \tag{10.53}$$

and hence, no noise reduction.

Using the matrix inversion lemma, (10.52) can be written as

$$\mathbf{w}_{\mathrm{SDW}} = \frac{\phi_{ss}\mathbf{\Phi_{vv}}^{-1}\mathbf{h}}{\mu + \phi_{ss}\mathbf{h}^H\mathbf{\Phi_{vv}}^{-1}\mathbf{h}}\, h_r^* \,, \tag{10.54}$$

and, using (10.9), the output SNR is equal to

$$\mathrm{SNR}_{\mathrm{SDW}}^{out} = \phi_{ss}\mathbf{h}^H\mathbf{\Phi_{vv}}^{-1}\mathbf{h} = \rho \,. \tag{10.55}$$

Implementation In practice, the speech correlation matrix $\phi_{ss}\mathbf{h}\mathbf{h}^H$ in (10.52) is unknown. During speech periods, the microphone signals consist of speech + noise, i.e., $x_m = h_m s + v_m$. During noise periods, only the noise component v_m is observed. The second-order statistics of the noise are assumed to be sufficiently stationary so that they can be estimated during periods of noise only. Assuming that the speech and the noise signal are uncorrelated, $\phi_{ss}\mathbf{h}\mathbf{h}^H$ can then be estimated as

$$\phi_{ss}\mathbf{h}\mathbf{h}^H \approx \mathbf{\Phi_{xx}} - \mathbf{\Phi_{vv}} \,, \tag{10.56}$$

where $\mathbf{\Phi_{xx}}$ is estimated during speech + noise periods and $\mathbf{\Phi_{vv}}$ during periods of noise only. Similarly as for the GSC, a robust speech detection is thus needed.

Different recursive implementations have been proposed for the SDW-MWF. In [29, 85], recursive matrix-decomposition-based implementations were presented,

which are computationally quite expensive. In [94], cheaper (time-domain and frequency-domain) stochastic gradient algorithms were proposed. These algorithms however require large circular data buffers, resulting in a large memory requirement. In [33], a frequency-domain criterion for the SDW-MWF was formulated, leading to frequency-domain stochastic gradient implementations using frequency-domain correlation matrices. The latter reduce the memory requirement and the computational complexity of the SDW-MWF.

Instead of estimating the speech correlation matrix $\phi_{ss}\mathbf{h}\mathbf{h}^H$ online using (10.56), it has been proposed to use a pre-determined estimate of $\phi_{ss}\mathbf{h}\mathbf{h}^H$ [43, 77]. In [77], this estimate is derived from clean speech recordings measured during an initial calibration phase. Additional recordings of the speech source signal allow to produce an estimate of the non-reverberant speech source signal instead of an estimate of the reverberant speech component in one of the microphone signals. However, since the room acoustics, the position of the desired speaker and the microphone characteristics may change over time, frequent re-calibration is required. In [43], a mathematical estimate of the correlation matrix and the correlation vector of the non-reverberant speech is exploited in which some signal model errors are taken into account.

Relationship between MWF and TF-GSC Simmer *et al.* [89] addressed the general problem of single-channel postfiltering. First, they derived the multi-channel Wiener filter for estimating the speech source signal s from the microphone signals $\mathbf{x}$ given in (10.3). Then, they showed that this Wiener filter can be factorised into a multiplication of the MVDR beamformer in (10.16) and a single-channel Wiener filter that depends on the speech and the noise components of the MVDR output. Balan and Rosca [3] showed that these results can be generalised to other multi-channel Bayesian cost functions, such as the short-time spectral amplitude (STSA) and the log spectral amplitude (LSA), proposed by Ephraim and Malah [37, 38].

Here, we generalise the results in [89] to the SDW-MWF estimating the speech component at the rth microphone. The SDW-MWF in (10.54) can be rewritten as

$$\mathbf{w}_{\text{SDW}} = \left(\frac{\phi_{ss}}{\phi_{ss} + \mu(\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h})^{-1}}\right)\frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}h_r^* \,. \tag{10.57}$$

Using $\mathbf{h} = h_r\tilde{\mathbf{h}}$ and $\phi_{ss} = \frac{\phi_{s_rs_r}}{|h_r|^2}$, we obtain

$$\mathbf{w}_{\text{SDW}} = \left(\frac{\phi_{s_rs_r}}{\phi_{s_rs_r} + \mu(\tilde{\mathbf{h}}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\tilde{\mathbf{h}})^{-1}}\right)\frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1}\tilde{\mathbf{h}}}{\tilde{\mathbf{h}}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\tilde{\mathbf{h}}} \,. \tag{10.58}$$

The reader can easily identify the second term in this expression as the modified MVDR beamformer $\tilde{\mathbf{w}}_{\text{MVDR}}$ in (10.37) which uses the RTFs rather than the ATFs. We now turn to the first term. Since by definition $y_s = s_r$, the speech PSD at the output of the modified MVDR beamformer is equal to

$$\phi_{y_sy_s} = \phi_{s_rs_r} \,. \tag{10.59}$$

The noise PSD at the output of the modified MVDR beamformer is equal to

$$\phi_{y_vy_v} = (\tilde{\mathbf{w}}_{\text{MVDR}})^H\mathbf{\Phi}_{vv}\tilde{\mathbf{w}}_{\text{MVDR}} = \left(\tilde{\mathbf{h}}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\tilde{\mathbf{h}}\right)^{-1}. \tag{10.60}$$

Using (10.59) and (10.60), the components of the first multiplicative term in the SDW-MWF can be identified as the speech distortion weighted single channel Wiener filter (SDW-SWF) applied to the modified MVDR beamformer output, i.e.

$$\mathbf{w}_{\text{SDW}} = \underbrace{\frac{\phi_{y_s y_s}}{\phi_{y_s y_s} + \mu \phi_{y_v y_v}}}_{\text{SDW-SWF postfilter}} \times \underbrace{\frac{\mathbf{\Phi}_{\mathbf{vv}}^{-1}\tilde{\mathbf{h}}}{\tilde{\mathbf{h}}^H \mathbf{\Phi}_{\mathbf{vv}}^{-1}\tilde{\mathbf{h}}}}_{\text{MVDR beamformer}} . \tag{10.61}$$

Mitigating the leakage problem of the GSC Due to room reverberation, microphone mismatch, look direction error and spatially distributed sources, speech components may leak into the noise references of the standard GSC, giving rise to speech distortion and possibly signal cancelation. To limit the distortion effect caused by the speech leakage, the ANC of the GSC is typically adapted during periods of noise only [45, 50, 54, 96]. When used in combination with small-sized arrays, e.g., in hearing aid applications, an additional robustness constraint [23, 52, 54, 56, 95] is required to guarantee performance in the presence of small errors in the assumed model, such as microphone mismatch [93]. A commonly used solution is to impose a quadratic inequality constraint (QIC) on the ANC filters $\mathbf{g}$, i.e.,

$$\mathbf{g}^H \mathbf{g} \leq \beta^2 . \tag{10.62}$$

The QIC avoids excessive growth of the ANC filters, hence reducing the undesired speech distortion when speech leaks into the noise references.

An alternative solution to limit speech distortion is to use an SDW-MWF instead of an ANC for the adaptive stage in the GSC, as depicted in Figure 10.4 [91]. Define the speech and the noise component of the noise reference signals of the GSC as $\mathbf{u}_s$ and $\mathbf{u}_v$, respectively, and define the speech and the noise component in the speech reference signal as $y_{s,MBF}$ and $y_{v,MBF}$, respectively. In contrast to the ANC that minimises the output noise power, i.e.

$$E\{|y_{v,MBF} - \mathbf{g}^H \mathbf{u}_v|^2\} , \tag{10.63}$$

the $(M-1)$-dimensional SDW-MWF $\tilde{\mathbf{w}}^H = \begin{bmatrix} \tilde{w}_2^* & \tilde{w}_3^* & \dots & \tilde{w}_M^* \end{bmatrix}$ takes speech distortion due to speech leakage explicitly into account in the design criterion, i.e.

$$\tilde{\mathbf{w}} = \underset{\tilde{\mathbf{w}}}{\text{argmin}} \frac{1}{\mu} E\{|\tilde{\mathbf{w}}^H \mathbf{u}_s|^2\} + E\{|y_{v,MBF} - \tilde{\mathbf{w}}^H \mathbf{u}_v|^2\} , \tag{10.64}$$

where the parameter $\mu \in [0, \infty]$ trades off noise reduction and speech distortion.

Since the SDW-MWF takes speech distortion explicitly into account in its design criterion, it is possible to include an extra filter $\tilde{w}_1$ on the speech reference (see Figure 10.4). Define the M-dimensional extended multi-channel filter $\tilde{\mathbf{w}}_e$ as

$$\tilde{\mathbf{w}}_e^H = \begin{bmatrix} \tilde{w}_1^* & \tilde{\mathbf{w}}^H \end{bmatrix} . \tag{10.65}$$

The optimisation criterion then becomes

$$\tilde{\mathbf{w}}_e = \underset{\tilde{\mathbf{w}}_e}{\text{argmin}} \frac{1}{\mu} E\left\{ \left| \tilde{\mathbf{w}}_e^H \begin{bmatrix} y_{s,MBF} \\ \mathbf{u_s} \end{bmatrix} \right|^2 \right\} + E\left\{ \left| y_{v,MBF} - \tilde{\mathbf{w}}_e^H \begin{bmatrix} y_{v,MBF} \\ \mathbf{u}_v \end{bmatrix} \right|^2 \right\} ,$$

Depending on the setting of the trade-off parameter μ and the presence/absence of the filter $\tilde{w}_1$ on the speech reference, different algorithms are obtained [91]:

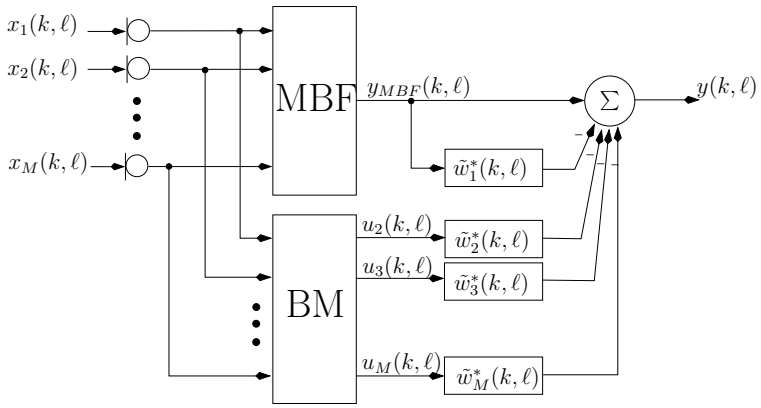


Figure 10.4 Application of the SDW-MWF to mitigate the speech leakage problem of the GSC.

- Without a filter $\tilde{w}_1$, we obtain the *speech distortion regularised GSC* (SDR-GSC), where the standard optimisation criterion of the ANC is supplemented with the regularisation term

$$\frac{1}{\mu} E\left\{|\tilde{\mathbf{w}}^H \mathbf{u}_s|^2\right\} . \tag{10.66}$$

  For $\mu = \infty$, speech distortion is completely ignored, which corresponds to the standard GSC. Compared to the QIC, the SDR-GSC is less conservative, since the regularisation term is proportional to the actual amount of speech leakage in the noise references. In [91, 94], it has been shown that in comparison with the QIC, the SDR-GSC achieves a better noise reduction performance for small model errors, while guaranteeing robustness against large model errors.

- With a filter $\tilde{w}_1$, we obtain the *spatially pre-processed speech distortion weighted multi-channel Wiener filter* (SP-SDW-MWF). In [91], it has been shown that for infinitely long filter lengths, the SP-SDW-MWF is not affected by microphone mismatch.

Illustration for a hearing aid application Figure 10.5 illustrates the noise reduction and speech distortion of the frequency-domain implementation of the SP-SDW-MWF for a hearing aid application [33]. In the implementation, a (non-perfect) energy-based voice activity detector (VAD) has been used [100]. A three-microphone behind-the-ear hearing aid was mounted on the right ear of a dummy head in an office room with a reverberation time $T_{60dB} \approx 700$ ms. The desired speech source and noise sources are positioned at a distance of 1 m from the head: the speech source in front of the head, the noise sources at an angle θ with respect to the speech source. The speech signal consists of sentences from the HINT-database [74] spoken by a male speaker. The noise scenario consists of 5 spectrally non-stationary multi-talker babble noise sources at 75°, 120°, 180°, 240°, 285°. The input SNR of the microphone signals is 0 dB. For evaluation purposes, the speech and noise signal were recorded separately.

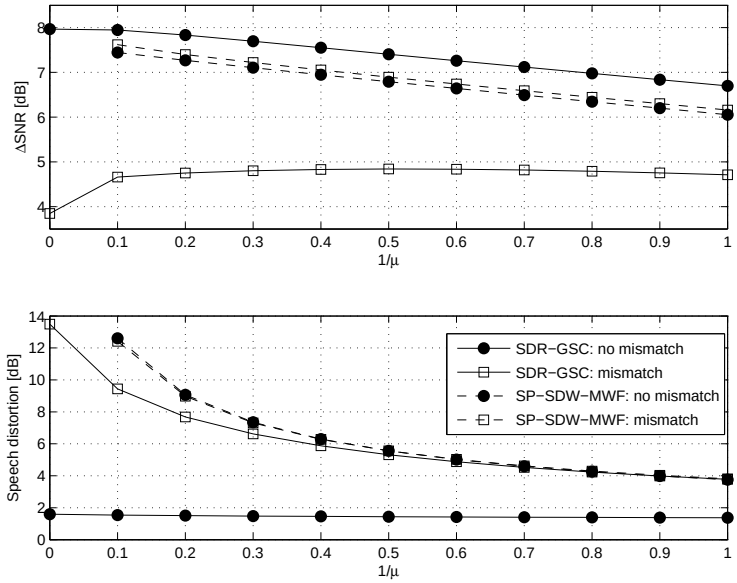


Figure 10.5 Illustration of the increased robustness of the SP-SDW-MWF for a hearing aid application.

To assess the performance, the (broadband) *intelligibility weighted SNR improvement* [44] between the output signal and the speech reference signal is computed, i.e.

$$\Delta\text{SNR} = \sum_{k=0}^{K-1} F(k) \left[10\log_{10}\text{SNR}^{out}(k) - 10\log_{10}\text{SNR}^{MBF}(k)\right] , \qquad (10.67)$$

where $F(k)$ expresses the importance of the kth frequency bin for speech intelligibility [80]. The intelligibility weighted speech distortion is defined similarly.

Figure 10.5 depicts the noise reduction and the speech distortion for the SDR-GSC and the SP-SDW-MWF as a function of $1/\mu$. This figure also demonstrates the effect of a gain mismatch of 4 dB applied to the second microphone.

- SDR-GSC: *In the absence of microphone mismatch*, the amount of speech leakage into the noise references is limited. Hence, the amount of speech distortion of the GSC (i.e., the SDR-GSC with $1/\mu = 0$) is low. Since there is still a small amount of speech leakage due to reverberation, the amount of noise reduction and speech distortion slightly decreases for increasing $1/\mu$. *In the presence of microphone mismatch*, the amount of speech leakage into the noise references grows. For $1/\mu = 0$ (GSC), the speech gets significantly distorted. Setting $1/\mu > 0$, improves the performance of the GSC in the presence of microphone mismatch, i.e., the speech distortion decreases and the degradation in noise reduction becomes smaller.

- SP-SDW-MWF: The noise reduction and speech distortion also decrease for increasing $1/\mu$. Compared to the SDR-GSC, the distortion is larger due to spectral filtering, but both the noise reduction and speech distortion are hardly affected by microphone mismatch.

10.4 BINAURAL BEAMFORMING

The main difference between monaural and binaural beamforming is the fact that we now consider two sets of microphones, i.e. one at the left hearing aid and one at the right hearing aid, and the fact that the binaural beamformer needs to produce two (distinct) output signals, cf. Figure 10.6. In this section we will consider a binaural system where both hearing aids are cooperating with each other, e.g. through a wireless link. Although currently available binaural hearing aids with a wireless link have very limited data rates and are only able to transmit data in order to coordinate parameter settings, we anticipate that future binaural hearing aids will also allow for transmission of (coded) audio signals [10], such that we may assume that all microphone signals are simultaneously available at the left and the right hearing aid. As already mentioned in Section 10.1, the objective of a binaural signal enhancement algorithm is not only to selectively extract the useful signal and to suppress the background noise, but also to preserve the binaural auditory cues (interaural time and level difference) of the speech and the noise sources, since these cues play a major role in the preservation of spatial awareness and in an improved speech understanding in noisy environments [6, 9, 12, 13, 48, 99].

Section 10.4.1 discusses the binaural microphone configuration and notations. Section 10.4.2 discusses the binaural MWF, where it is shown that the binaural cues of the speech source are preserved, but the binaural cues of the noise component are distorted. Sections 10.4.3 and 10.4.4 then discuss two extensions of the binaural MWF that aim to also preserve the binaural cues of the noise component. In Section 10.4.3 a parameter allows to trade off noise reduction performance and preservation of the binaural noise cues, whereas in Section 10.4.4 the binaural MWF cost function is extended with terms related to the interaural transfer function (ITF). Section 10.4.5 presents experimental results for the discussed binaural noise reduction algorithms.

10.4.1 Configuration and notation

Microphone signals and output signals Similarly to the monaural configuration in Figure 10.1, the binaural hearing aid configuration in Figure 10.6 now consists of two microphone arrays, each having M microphones. The mth microphone signal on the left hearing aid $x_{L,m}$ can be written as

$$x_{L,m} = h_{L,m}s + v_{L,m}, \quad m = 1 \dots M,$$

with $h_{L,m}$ the mth element of the ATF for the left hearing aid. Similarly, the mth microphone signal on the right hearing aid is equal to $x_{R,m} = h_{R,m}s + v_{R,m}$. Hence, the signal model in (10.3) remains valid, i.e.

$$\mathbf{x} = \mathbf{h}s + \mathbf{v}\,, \tag{10.68}$$

where $\mathbf{x}$, $\mathbf{v}$ and $\mathbf{h}$ are now $2M$-dimensional vectors, defined as

$$\mathbf{x} = \begin{bmatrix} \mathbf{x}_L \\ \mathbf{x}_R \end{bmatrix}, \quad \mathbf{v} = \begin{bmatrix} \mathbf{v}_L \\ \mathbf{v}_R \end{bmatrix}, \quad \mathbf{h} = \begin{bmatrix} \mathbf{h}_L \\ \mathbf{h}_R \end{bmatrix}\,, \tag{10.69}$$

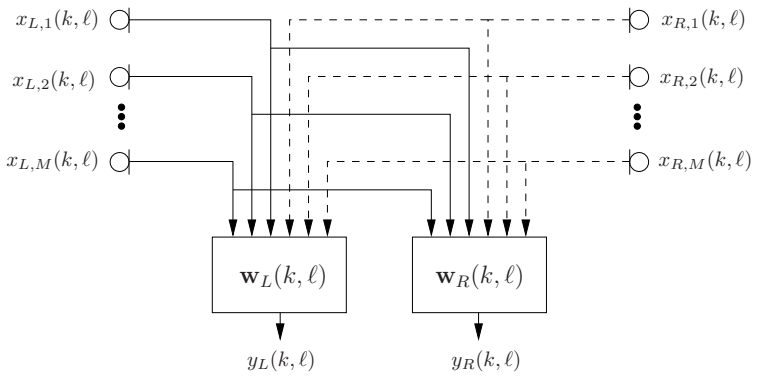


Figure 10.6 General binaural processing scheme

with

$$\mathbf{x}_L = \begin{bmatrix} x_{L,1} & x_{L,2} & \cdots & x_{L,M} \end{bmatrix}^T \qquad \mathbf{h}_L = \begin{bmatrix} h_{L,1} & h_{L,2} & \cdots & h_{L,M} \end{bmatrix}^T ,$$
$$\mathbf{x}_R = \begin{bmatrix} x_{R,1} & x_{R,2} & \cdots & x_{R,M} \end{bmatrix}^T \qquad \mathbf{h}_R = \begin{bmatrix} h_{R,1} & h_{R,2} & \cdots & h_{R,M} \end{bmatrix}^T .$$

We will use the r_L-th microphone on the left hearing aid and the r_R-th microphone on the right hearing aid as the reference microphones for the speech enhancement algorithms. Typically, the front microphones are used as reference microphones. For conciseness, the reference microphone signals x_{L,r_L} and x_{R,r_R} on the left and the right hearing aid are denoted as x_L and x_R, i.e.

$$x_L = \mathbf{e}_L^H \mathbf{x} = h_L s + v_L, \quad x_R = \mathbf{e}_R^H \mathbf{x} = h_R s + v_R , \tag{10.70}$$

where $\mathbf{e}_L$ and $\mathbf{e}_R$ are $2M$-dimensional canonical vectors, cf. (10.53), with $\mathbf{e}_L(r_L) = 1$ and $\mathbf{e}_R(M + r_R) = 1$.

The output signals y_L and y_R on the left and the right hearing aid are obtained by filtering and summing all microphone signals from both hearing aids, i.e.

$$y_L = \mathbf{w}_L^H \mathbf{x}, \quad y_R = \mathbf{w}_R^H \mathbf{x} , \tag{10.71}$$

where $\mathbf{w}_L$ and $\mathbf{w}_R$ represent the $2M$-dimensional beamformer filter coefficients for the left and the right hearing aid. Similarly as in (10.8), the output signals can be decomposed as

$$y_L = \mathbf{w}_L^H \mathbf{h} s + \mathbf{w}_L^H \mathbf{v} \triangleq y_{sL} + y_{vL} , \tag{10.72}$$
$$y_R = \mathbf{w}_R^H \mathbf{h} s + \mathbf{w}_R^H \mathbf{v} \triangleq y_{sR} + y_{vR} . \tag{10.73}$$

The $4M$-dimensional stacked weight vector $\mathbf{w}$ is defined as

$$\mathbf{w} = \begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix} . \tag{10.74}$$

Performance measures Similarly as in (10.9), the *SNR* of the beamformer output on the left and the right hearing aid is defined as

$$\mathrm{SNR}_L^{out} = \frac{E\{|y_{sL}|^2\}}{E\{|y_{vL}|^2\}} = \frac{|\mathbf{w}_L^H\mathbf{h}|^2\phi_{ss}}{\mathbf{w}_L^H\mathbf{\Phi_{vv}}\mathbf{w}_L}, \quad \mathrm{SNR}_R^{out} = \frac{E\{|y_{sR}|^2\}}{E\{|y_{vR}|^2\}} = \frac{|\mathbf{w}_R^H\mathbf{h}|^2\phi_{ss}}{\mathbf{w}_R^H\mathbf{\Phi_{vv}}\mathbf{w}_R} . \tag{10.75}$$

Since for binaural beamformers we are not only concerned with noise reduction but also with binaural cue preservation, we define the ILD and the ITD of the speech and the noise components, both on the (input) reference microphones and on the beamformer output. The *ILD* for the speech and the noise component is expressed as the power ratio between the left and the right hearing aid, i.e.

$$P_v^{in} = \frac{E\{|v_L|^2\}}{E\{|v_R|^2\}} = \frac{\mathbf{e}_L^H\mathbf{\Phi_{vv}}\mathbf{e}_L}{\mathbf{e}_R^H\mathbf{\Phi_{vv}}\mathbf{e}_R} , \quad P_v^{out} = \frac{E\{|y_{vL}|^2\}}{E\{|y_{vR}|^2\}} = \frac{\mathbf{w}_L^H\mathbf{\Phi_{vv}}\mathbf{w}_L}{\mathbf{w}_R^H\mathbf{\Phi_{vv}}\mathbf{w}_R} , \tag{10.76}$$

$$P_s^{in} = \frac{|h_L|^2}{|h_R|^2} , \qquad P_s^{out} = \frac{E\{|y_{sL}|^2\}}{E\{|y_{sR}|^2\}} = \frac{|\mathbf{w}_L^H\mathbf{h}|^2}{|\mathbf{w}_R^H\mathbf{h}|^2} . \tag{10.77}$$

The *ITD* is expressed as the (phase of) the cross-correlation, i.e.

$$c_v^{in} = E\{v_L v_R^*\} = \mathbf{e}_L^H\mathbf{\Phi_{vv}}\mathbf{e}_R , \qquad c_v^{out} = E\{y_{vL}y_{vR}^*\} = \mathbf{w}_L^H\mathbf{\Phi_{vv}}\mathbf{w}_R , \tag{10.78}$$

$$c_s^{in} = \phi_{ss}h_L h_R^* , \qquad c_s^{out} = E\{y_{sL}y_{sR}^*\} = \phi_{ss}\mathbf{w}_L^H\mathbf{h}\mathbf{h}^H\mathbf{w}_R . \tag{10.79}$$

Combining ITD and ILD[4], the *interaural transfer function (ITF)* is defined as the ratio of the components at the left and the right hearing aid, i.e.

$$ITF_v^{in} = \frac{v_L}{v_R} , \qquad ITF_v^{out} = \frac{y_{vL}}{y_{vR}} = \frac{\mathbf{w}_L^H\mathbf{v}}{\mathbf{w}_R^H\mathbf{v}} , \tag{10.80}$$

$$ITF_s^{in} = \frac{h_L}{h_R} , \qquad ITF_s^{out} = \frac{y_{sL}}{y_{sR}} = \frac{\mathbf{w}_L^H\mathbf{h}}{\mathbf{w}_R^H\mathbf{h}} . \tag{10.81}$$

10.4.2 Binaural multi-channel Wiener Filter

Cost function Similarly as the monaural MWF discussed in Section 10.3.5, the binaural MWF produces an MMSE estimate of the speech component on both hearing aids [63]. The MSE cost function for the filter $\mathbf{w}_L$ estimating the speech component $s_L = h_L s$ in the left reference microphone signal and the filter $\mathbf{w}_R$ estimating the speech component $s_R = h_R s$ in the right reference microphone signal is equal to

$$J_{MSE}(\mathbf{w}) = E\left\{\left\|\begin{bmatrix} s_L - y_L \\ s_R - y_R \end{bmatrix}\right\|^2\right\} = E\left\{\left\|\begin{bmatrix} s_L - \mathbf{w}_L^H\mathbf{x} \\ s_R - \mathbf{w}_R^H\mathbf{x} \end{bmatrix}\right\|^2\right\} . \tag{10.82}$$

The cost function for the binaural SDW-MWF, allowing a trade-off between speech distortion and noise reduction, is equal to

$$J_{SDW}(\mathbf{w}) = E\left\{\left\|\begin{bmatrix} s_L - \mathbf{w}_L^H\mathbf{h}s \\ s_R - \mathbf{w}_R^H\mathbf{h}s \end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix} \mathbf{w}_L^H\mathbf{v} \\ \mathbf{w}_R^H\mathbf{v} \end{bmatrix}\right\|^2\right\} , \tag{10.83}$$

[4]It can be easily verified that the phase of the ITF is equal to the ITD, and the squared norm of the ITF is equal to the ILD.

where μ trades off noise reduction and speech distortion. Similarly as (10.52) and (10.54) for the monaural case, the binaural SDW-MWF solution is equal to [21, 28]

$$\mathbf{w}_{SDW,L} = \left(\phi_{ss}\mathbf{h}\mathbf{h}^H + \mu\mathbf{\Phi}_{\mathbf{vv}}\right)^{-1}\phi_{ss}\mathbf{h}h_L^* = \frac{\phi_{ss}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{\mu+\rho}h_L^*, \quad (10.84)$$

$$\mathbf{w}_{SDW,R} = \left(\phi_{ss}\mathbf{h}\mathbf{h}^H + \mu\mathbf{\Phi}_{\mathbf{vv}}\right)^{-1}\phi_{ss}\mathbf{h}h_R^* = \frac{\phi_{ss}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{\mu+\rho}h_R^*\,, \quad (10.85)$$

with $\rho = \phi_{ss}\mathbf{h}^H\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}$. Hence, it can be shown that the output SNR at both hearing aids is identical and equal to ρ.

Cue preservation From (10.84) and (10.85), it can be seen that the binaural MWF vectors for the left and the right hearing aid are parallel, i.e.

$$\mathbf{w}_{SDW,L} = ITF_s^{in,*}\,\mathbf{w}_{SDW,R}\,, \quad (10.86)$$

with ITF_s^{in} defined in (10.81). Hence, the ITF of the output speech and noise components are both equal to ITF_s^{in}, implying that the binaural speech cues are perfectly preserved, but the binaural noise cues are distorted. As all output components are perceived as coming from the speech direction, the auditory perception of the acoustic scene is therefore typically not preserved by the binaural MWF.

10.4.3 Partial noise estimation (SDW-MWF-η)

Cost function An extension of the binaural MWF in (10.82) that aims to preserve the binaural noise cues (in addition to preserving the binaural speech cues) has been proposed in [63], where the objective is to produce an MMSE estimate of the sum of the speech component and a scaled version of the noise component in the reference microphone signals, i.e.

$$J_{MSE\eta}(\mathbf{w}) = E\left\{\left\|\begin{bmatrix}(s_L+\eta v_L)-\mathbf{w}_L^H\mathbf{x}\\(s_R+\eta v_R)-\mathbf{w}_R^H\mathbf{x}\end{bmatrix}\right\|^2\right\}, \quad (10.87)$$

with the scaling parameter $0 \leq \eta \leq 1$. When $\eta = 0$, this cost function reduces to the standard binaural MWF cost function in (10.82). When $\eta = 1$, the optimal filters are obviously equal to $\mathbf{w}_{MSE\eta,L} = \mathbf{e}_L$ and $\mathbf{w}_{MSE\eta,R} = \mathbf{e}_R$, resulting in perfect preservation of the binaural speech and noise cues, but no noise reduction at all. This partial noise estimation procedure is in fact an extension of single-channel Wiener filters that have been presented in [5, 25].

Similarly to the SDW-MWF cost function in (10.83), it is also possible to combine partial noise estimation with trading off noise reduction and speech distortion, i.e.

$$J_{SDW\eta}(\mathbf{w}) = E\left\{\left\|\begin{bmatrix}s_L-\mathbf{w}_L^H\mathbf{h}s\\s_R-\mathbf{w}_R^H\mathbf{h}s\end{bmatrix}\right\|^2 + \mu\left\|\begin{bmatrix}\eta v_L-\mathbf{w}_L^H\mathbf{v}\\\eta v_R-\mathbf{w}_R^H\mathbf{v}\end{bmatrix}\right\|^2\right\}. \quad (10.88)$$

The binaural filters minimising $J_{SDW\eta}(\mathbf{w})$ are equal to

$$\begin{aligned}\mathbf{w}_{SDW\eta,L} &= (\phi_{ss}\mathbf{h}\mathbf{h}^H + \mu\mathbf{\Phi}_{\mathbf{vv}})^{-1}(\phi_{ss}\mathbf{h}\mathbf{h}^H + \eta\mu\mathbf{\Phi}_{\mathbf{vv}})\mathbf{e}_L \\ &= (1-\eta)\mathbf{w}_{SDW,L} + \eta\mathbf{e}_L = (1-\eta)\frac{\phi_{ss}\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}}{\mu+\rho}h_L^* + \eta\mathbf{e}_L\,, \quad (10.89)\end{aligned}$$

$$\begin{aligned}\mathbf{w}_{SDW\eta,R} &= (\phi_{ss}\mathbf{h}\mathbf{h}^H + \mu\mathbf{\Phi_{vv}})^{-1}(\phi_{ss}\mathbf{h}\mathbf{h}^H + \eta\mu\mathbf{\Phi_{vv}})\mathbf{e}_R \\ &= (1-\eta)\mathbf{w}_{SDW,R} + \eta\mathbf{e}_R = (1-\eta)\frac{\phi_{ss}\mathbf{\Phi_{vv}}^{-1}\mathbf{h}}{\mu+\rho}h_R^* + \eta\mathbf{e}_R \ . \quad (10.90)\end{aligned}$$

Hence, the SDW-MWF with partial noise estimation corresponds to mixing the output signal of the standard SDW-MWF with the reference microphone signals.

It has been shown in [21] that the SNR improvement at the right hearing aid is equal to

$$\Delta SNR_R = \Delta SNR_R^o \frac{\left(\frac{\eta\mu+\rho}{\mu+\rho}\right)^2}{\left(\frac{\eta\mu+\rho}{\mu+\rho}\right)^2 + (\Delta SNR_R^o - 1)\eta^2} \ , \quad (10.91)$$

where ΔSNR_R^o represents the SNR improvement at the right hearing aid for the standard binaural SDW-MWF in (10.85). Obviously, if $\eta = 0$, the SNR improvement is equal to ΔSNR_R^o, whereas if $\eta = 1$, no SNR improvement is obtained, i.e. $\Delta SNR_R = 1$. Since the SNR improvement ΔSNR_R^o of the standard SDW-MWF is always larger than or equal to 1 [32], it can be easily shown that

$$1 \leq \Delta SNR_R \leq \Delta SNR_R^o \ , \quad (10.92)$$

i.e. the SNR improvement is always smaller when using partial noise estimation. Similar expressions can be derived for the left hearing aid, where it should also be noted that the output SNR at both hearing aids is now not necessarily identical.

Cue preservation From (10.89) and (10.90), it can be seen that the binaural MWF filters when using partial noise estimation are in general not parallel, such that the ITF of the output speech and noise components are typically different. Using (10.81), the ITF of the output speech component is equal to

$$ITF_s^{out} = \frac{\mathbf{w}_{SDW\eta,L}^H\mathbf{h}}{\mathbf{w}_{SDW\eta,R}^H\mathbf{h}} = \frac{(1-\eta)\frac{\rho}{\mu+\rho}h_L + \eta h_L}{(1-\eta)\frac{\rho}{\mu+\rho}h_R + \eta h_R} = \frac{h_L}{h_R} = ITF_s^{in} \ , \quad (10.93)$$

such that the binaural speech cues are preserved. Using (10.76) and (10.77), it can be shown that the ILD of the output noise component is a weighted sum of the ILDs of the input speech and noise component [21], i.e.

$$P_v^{out} = P_s^{in} + \frac{\eta^2}{\psi\ SNR_R^{in} + \eta^2}\ (P_v^{in} - P_s^{in}) \ , \quad (10.94)$$

with

$$\psi = \frac{(1-\eta)(2\eta\mu + \rho + \eta\rho)}{(\mu+\rho)^2} \ , \quad (10.95)$$

and SNR_R^{in} representing the input SNR on the right hearing aid. If $\eta = 0$, the ILD of the output noise component is equal to P_s^{in}, whereas if $\eta = 1$, ψ is equal to 0 such that the ILD of the output noise component is equal to P_v^{in}. Similarly, using (10.78) and (10.79), it can be shown that the cross-correlation (related to the ITD) of the output noise component is a weighted sum of the cross-correlations of the input speech and noise component [21], i.e.

$$c_v^{out} = \psi\ c_s^{in} + \eta^2 c_v^{in} \ . \quad (10.96)$$

Hence, the parameter η trades off noise reduction and preservation of the binaural noise cues, as will also be shown in the experimental results in Section 10.4.5.

10.4.4 ITF cost function (SDW-MWF-ITF)

Cost function In order to control the binaural cues of the speech and the noise component, it is also possible to extend the quadratic SDW-MWF cost function in (10.83) with terms related to the ITF of the speech and the noise components, as has been proposed in [28, 62, 97]. Since the aim is to preserve the binaural speech and noise cues, the desired output ITFs ITF_v^{des} and ITF_s^{des} are equal to the input ITFs defined in (10.80) and (10.81), where for the noise component the input ITF is assumed to be constant (as is e.g. the case for a single noise source), such that it can be estimated in the least squares sense using the noise correlation matrix, i.e.

$$ITF_v^{des} = \frac{E\{v_L v_R^*\}}{E\{v_R v_R^*\}} = \frac{\mathbf{e}_L^H \mathbf{\Phi_{vv}} \mathbf{e}_R}{\mathbf{e}_R^H \mathbf{\Phi_{vv}} \mathbf{e}_R}, \quad ITF_s^{des} = \frac{h_L}{h_R} \,. \tag{10.97}$$

The (non-linear) *ITF cost function* for preserving the binaural cues of the noise component is then defined as [28]

$$J_{ITF,1}^v(\mathbf{w}) = E\Big\{ \Big| \frac{\mathbf{w}_L^H \mathbf{v}}{\mathbf{w}_R^H \mathbf{v}} - ITF_v^{des} \Big|^2 \Big\} \,, \tag{10.98}$$

which, in the case of a single noise source, is equal to

$$J_{ITF,1}^v(\mathbf{w}) = \frac{E\{|\mathbf{w}_L^H \mathbf{v} - ITF_v^{des}\, \mathbf{w}_R^H \mathbf{v}|^2\}}{E\{|\mathbf{w}_R^H \mathbf{v}|^2\}} = \frac{\mathbf{w}^H \mathbf{\Phi}_{\mathbf{vv}}^t \mathbf{w}}{\mathbf{w}^H \mathbf{\Phi}_{\mathbf{vv}}^1 \mathbf{w}} \,, \tag{10.99}$$

with

$$\mathbf{\Phi}_{\mathbf{vv}}^t = \begin{bmatrix} \mathbf{\Phi_{vv}} & -ITF_v^{des,*}\, \mathbf{\Phi_{vv}} \\ -ITF_v^{des}\, \mathbf{\Phi_{vv}} & |ITF_v^{des}|^2\, \mathbf{\Phi_{vv}} \end{bmatrix} \,, \quad \mathbf{\Phi}_{\mathbf{vv}}^1 = \begin{bmatrix} \mathbf{0}_{2M} & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \mathbf{\Phi_{vv}} \end{bmatrix} \,.$$

We also introduce a simplified *quadratic ITF cost function* by assuming the denominator in (10.99) to be constant, i.e. [62, 97]

$$J_{ITF,2}^v(\mathbf{w}) = E\{|\mathbf{w}_L^H \mathbf{v} - ITF_v^{des}\, \mathbf{w}_R^H \mathbf{v}|^2\} = \mathbf{w}^H \mathbf{\Phi}_{\mathbf{vv}}^t \mathbf{w} \,. \tag{10.100}$$

The ITF cost function for the speech component is defined similarly as the ITF cost function for the noise component, by replacing the noise correlation matrix $\mathbf{\Phi_{vv}}$ with the speech correlation matrix $\phi_{ss}\mathbf{h}\mathbf{h}^H$ and the desired noise ITF with the desired speech ITF. The *total cost function* trading off noise reduction, speech distortion and binaural cue preservation is then defined as

$$J_{SDW-ITF}(\mathbf{w}) = J_{SDW}(\mathbf{w}) + \gamma J_{ITF}^s(\mathbf{w}) + \delta J_{ITF}^v(\mathbf{w}) \,, \tag{10.101}$$

where the parameters γ and δ enable to put more emphasis on binaural cue preservation for the speech and the noise component. When using the non-linear ITF cost function in (10.99), no closed-form expression is available for the filter minimising $J_{SDW-ITF}(\mathbf{w})$, such that we have to use iterative optimisation techniques [28]. On the other hand, when using the quadratic ITF cost function in (10.100), the filter minimising $J_{SDW-ITF}(\mathbf{w})$ is equal to

$$\mathbf{w}_{SDW-ITF} = (\mathbf{\Phi} + \gamma\delta \mathbf{\Phi}_{\mathbf{ss}}^t + \delta \mathbf{\Phi}_{\mathbf{vv}}^t)^{-1} \mathbf{r}_x \,, \tag{10.102}$$

with

$$\mathbf{\Phi} = \begin{bmatrix} \phi_{ss}\mathbf{h}\mathbf{h}^H + \mu\mathbf{\Phi}_{\mathbf{vv}} & \mathbf{0}_{2M} \\ \mathbf{0}_{2M} & \phi_{ss}\mathbf{h}\mathbf{h}^H + \mu\mathbf{\Phi}_{\mathbf{vv}} \end{bmatrix}, \quad \mathbf{r}_x = \begin{bmatrix} \phi_{ss}\mathbf{h}h_L^* \\ \phi_{ss}\mathbf{h}h_R^* \end{bmatrix}. \tag{10.103}$$

For $\gamma = 0$, it can be shown that the filter $\mathbf{w}_{SDW-ITF}$ is equal to[5] [20, 21]

$$\mathbf{w}_{SDW-ITF,L} = \frac{\phi_{ss}}{\mu+\rho}\left[h_L^* - \xi(h_L^* - ITF_v^{des,*}h_R^*)\right]\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}\,, \tag{10.104}$$

$$\mathbf{w}_{SDW-ITF,R} = \frac{\phi_{ss}}{\mu+\rho}\left[h_R^* + \xi ITF_v^{des}(h_L^* - ITF_v^{des,*}h_R^*)\right]\mathbf{\Phi}_{\mathbf{vv}}^{-1}\mathbf{h}\,, \tag{10.105}$$

with

$$\xi = \frac{\delta}{\mu+\rho+\delta(1+|ITF_v^{des}|^2)}\,. \tag{10.106}$$

Hence, the filter $\mathbf{w}_{SDW-ITF}$ is equal to $\mathbf{w}_{SDW}$ in (10.84)-(10.85) plus an extra term due to the ITF extension. Remarkably, it can be shown that the output SNR at both hearing aids is again identical and equal to ρ.

Cue preservation From (10.104) and (10.105), it follows that the filters $\mathbf{w}_{SDW-ITF,L}$ and $\mathbf{w}_{SDW-ITF,R}$ are parallel, such that the ITF of the output speech and noise components are identical and equal to

$$ITF^{out} = \frac{ITF_s^{in} - \xi(ITF_s^{in} - ITF_v^{des})}{1+\xi ITF_v^{des,*}(ITF_s^{in} - ITF_v^{des})}\,. \tag{10.107}$$

This however implies that typically the binaural cues of neither the speech nor the noise component are perfectly preserved. If the weight $\delta = 0$, then $\xi = 0$, such that the output ITF is equal to the input ITF of the speech component. On the other hand, if $\delta \to \infty$, then $\xi = 1/(1+|ITF_v^{des}|^2)$, such that the output ITF is equal to ITF_v^{des}, i.e. the desired ITF for the noise component. It is important to note that the output ITF in (10.107) is also linked to the output SNR ρ through the factor ξ. Hence, for frequencies with a large output SNR, i.e. large ρ and small ξ, the output ITF is close to the ITF of the speech component, whereas for frequencies with a small output SNR the output ITF is close to the ITF of the noise component [20, 21]. This advantageous perceptual effect has been observed in [97].

10.4.5 Experimental results

Set-up and performance measures Two hearing aids with $M = 2$ omni-directional microphones each have been mounted on a dummy head in a room with a reverberation time $T_{60dB} \approx 140$ ms. The distance between the microphones on each hearing aid is approximately 1 cm. The desired speech source and the noise sources are positioned at a distance of 1 m from the head, and different configurations have been considered:

- speech source in front of the head ($\theta_s = 0°$) and several noise configurations θ_v : a single noise source at 60°, 90°, 120°, 180°, 270° or 300°; two noise

[5] Similar expressions can be derived when $\gamma \neq 0$ [21].

sources at [−60° 60°], [−120° 120°], or [120° 210°]; and 4 noise sources at [60° 120° 180° 210°] (condition N4a) or [60° 120° 180° 270°] (condition N4b);

- speech source at the right side ($\theta_s = 90°$) and a noise source at $\theta_v = 180°$.

The speech signal consists of sentences from the HINT database [74], and the noise source is multi-talker babble noise. For all configurations, the input broadband SNR is 0 dB in the front microphone signal on the left hearing aid. For evaluation purposes, the speech and noise signal were recorded separately.

The binaural noise reduction techniques presented in Sections 10.4.2-10.4.4 are applied in each frequency bin. Using a perfect voice activity detector (VAD), the noise correlation matrix $\mathbf{\Phi_{vv}}$ in each frequency bin is computed during noise-only periods, the correlation matrix $\mathbf{\Phi_{xx}}$ in each frequency bin is computed during speech-and-noise periods, and similarly as for the monaural case in (10.56), the speech correlation matrix is estimated as $\phi_{ss}\mathbf{h}\mathbf{h}^H \approx \mathbf{\Phi_{xx}} - \mathbf{\Phi_{vv}}$. The FFT size used for STFT-processing is equal to $K = 128$ and the trade-off parameter $\mu = 5$.

To assess the performance of the different algorithms, the broadband intelligibility weighted SNR improvement between the output signal and the reference microphone signal is computed for the left and the right hearing aid, similarly as in (10.67). Since it is known that the ITD cue only plays a role in sound localisation at low frequencies [9, 13, 48], the *broadband ITD error* is computed as the sum of the ITD errors for frequencies below 1500 Hz, i.e. for the noise component

$$\Delta\text{ITD}_v^b = \sum_{k=0}^{K_{ITD}} |\angle c_v^{out}(k) - \angle c_v^{in}(k)| \,. \tag{10.108}$$

On the other hand, since ILD errors at each frequency are able to contribute to incorrect localisation, the *broadband ILD error* is computed as the sum of the ILD errors at each frequency (without weighting), i.e. for the noise component

$$\Delta\text{ILD}_v^b = \sum_{k=0}^{K-1} \left[10\log_{10}\text{P}_v^{out}(k) - 10\log_{10}\text{P}_v^{in}(k)\right] \,. \tag{10.109}$$

Performance comparison For different speech and noise configurations, Figure 10.7 depicts the SNR improvement at the left and the right hearing aid for 3 noise reduction algorithms: (a) a bilateral MWF, i.e. two independent MWF algorithms on the left and the right hearing aid using 2 microphones; (b) a binaural MWF using 2 microphones on the left (right) hearing aid and the front microphone on the right (left) hearing aid, (c) a binaural MWF using all 4 microphones from both hearing aids. In general, for all algorithms the SNR improvement is larger when the speech source and the noise source(s) are spatially more separated, with the largest SNR improvement occurring in the hearing aid where the input SNR is lower, e.g. for $\theta_v = 60°$ the SNR improvement is larger in the right hearing aid than in the left hearing aid. Not surprisingly, using more microphones leads to a larger SNR improvement, and this is even more pronounced when the speech source and the noise source(s) are close to each other (e.g. $\theta_v = 60°$) or when multiple noise sources are present.

For the configuration $\theta_s = 0°$ and $\theta_v = 60°$, Figure 10.8 depicts the SNR improvement (left and right hearing aid) and the ILD and ITD error (speech and noise component) for 2 binaural noise reduction algorithms using all 4 microphones: (a) the partial noise estimation SDW-MWF-η, as a function of the parameter η, (b) the SDW-MWF-ITF, as a function of the parameter δ (we assume $\gamma = 0$). First,

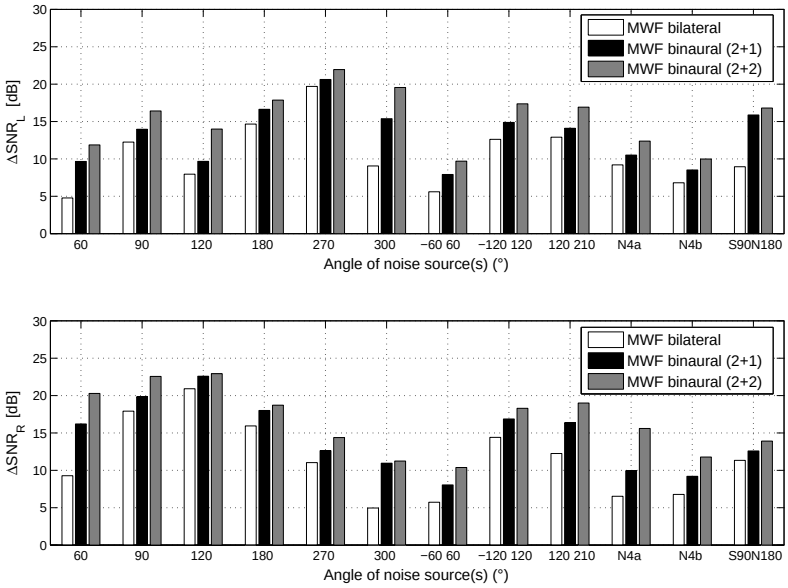


Figure 10.7 SNR improvement at left and right hearing aid for different speech and noise configurations for bilateral MWF (independent processing), and binaural MWF using 3 and 4 microphones

it can be observed that for the standard binaural MWF, i.e. $\eta = 0$ or $\delta = 0$, the ILD and ITD errors for the speech component are very small, whereas the ILD and ITD errors for the noise component are large, i.e. the binaural speech cues are preserved and the binaural noise cues are distorted. When increasing η for the SDW-MWF-η, the ILD and ITD errors for the noise component decrease, but the SNR improvement also decreases substantially. However, perceptual experiments have shown that e.g. for $\eta = 0.2$ it is possible to improve the sound localisation performance while hardly reducing the speech intelligibility [98]. When increasing δ for the SDW-MWF-ITF, the ILD and ITD errors for the noise component also decrease and the effect on the SNR improvement is smaller than for the SDW-MWF-η, but the ILD and ITD errors for the speech component increase. However, initial perceptual experiments still show an advantageous effect [97], which may be due to the frequency dependence of the output ITF on the output SNR, as explained in Section 10.4.4.

10.5 CONCLUSION

In this chapter we reviewed several multi-microphone beamforming techniques that can be used for noise reduction in monaural and binaural hearing aids. For monaural noise reduction we focused on the transfer-function GSC and the multi-channel Wiener filter, which are derived from different cost functions but are in fact closely related. For binaural noise reduction we presented several extensions of the standard binaural MWF that aim to preserve the binaural cues of the speech and the noise sources.

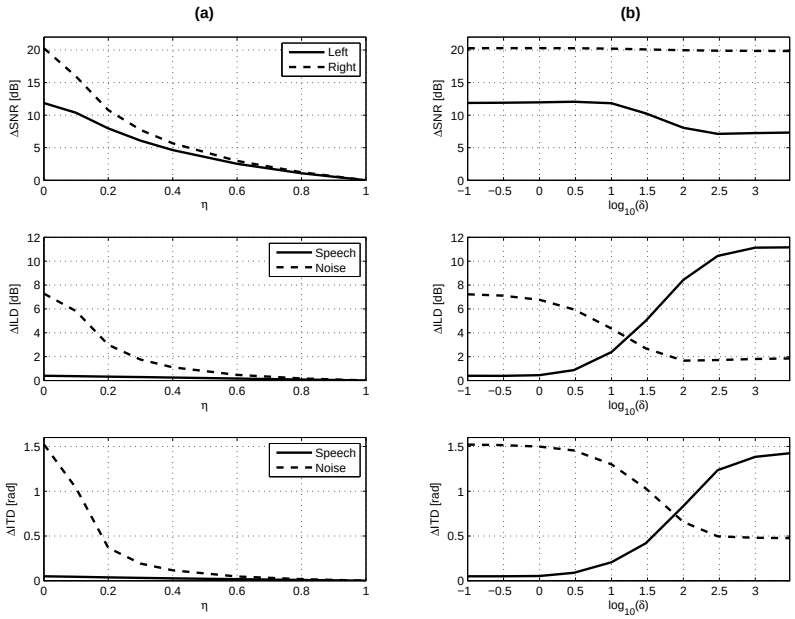


Figure 10.8 SNR improvement and ILD and ILD error for **(a)** SDW-MWF-η and **(b)** SDW-MWF-ITF ($\theta_s = 0^\circ$, $\theta_v = 60^\circ$)

REFERENCES

1. R. Aichner, W. Herbordt, H. Buchner, and W. Kellermann. Least-squares error beamforming using minimum statistics and multichannel frequency-domain adaptive filtering. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 223–226, Kyoto, Japan, September 2003.
2. M. R. Bai and C. Lin. Microphone array signal processing with application in three-dimensional hearing. *Journal of the Acoustical Society of America*, 117(4):2112–2121, April 2005.
3. R. Balan and J. Rosca. Microphone array speech enhancement by Bayesian estimation of spectral amplitude and phase. *Proc. Sensor Array and Multichannel Signal Processing Workshop*, pages 209–213, August 2002.
4. A. J. Bell and T. J. Sejnowski. An information-maximisation approach to blind separation and blind deconvolution. *Neural Computation*, 7(6):1004–1034, 1995.
5. J. Benesty, J. Chen, A. Huang, and S. Doclo. New Insights into the Noise Reduction Wiener Filter. *IEEE Trans. Audio, Speech and Language Processing*, 14(4):1218–1234, July 2006.
6. R. Beutelmann and T. Brand. Prediction of speech intelligibility in spatial noise and reverberation for normal-hearing and hearing-impaired listeners. *Journal of the Acoustical Society of America*, 120(1):331–342, July 2006.
7. J. Bitzer, K.-D. Kammeyer, and K. U. Simmer. An Alternative Implementation of the Superdirective Beamformer. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 7–10, New Paltz NY, USA, October 1999.

8. J. Bitzer and K. U. Simmer. *Superdirective Microphone Arrays*, chapter 2 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 19–38. Springer-Verlag, May 2001.

9. J. Blauert. *Spatial Hearing: The Psychophysics of Human Sound Localisation*. MIT Press, Cambridge MA, USA, 1983.

10. A. Boothroyd, K. Fitz, J. Kindred, S. Kochkin, H. Levitt, B.C.J. Moore, and J. Yantz. Hearing aids and wireless technology. *Hearing Review*, 14(6):44–48, June 2007.

11. B. R. Breed and J. Strauss. A short proof of the equivalence of LCMV and GSC beamforming. *IEEE Signal Processing Lett.*, 9(6):168–169, June 2002.

12. A. S. Bregman. *Auditory Scene Analysis*. MIT Press, Cambridge MA, USA, 1990.

13. A.W. Bronkhorst and R. Plomp. The effect of head-induced interaural time and level differences on speech intelligibility in noise. *Journal of the Acoustical Society of America*, 83(4):1508–1516, April 1988.

14. G. J. Brown and D. L. Wang. *Speech Enhancement*, chapter Separation of speech by computational auditory scene analysis, pages 371–402. Springer-Verlag, 2005.

15. H. Buchner, R. Aichner, and W. Kellermann. A generalization of Blind Source Separation algorithms for convolutive mixtures based on second-order statistics. *IEEE Trans. Speech and Audio Processing*, 13(1):120–134, January 2005.

16. J. Capon. High resolution frequency-wavenumber spectrum analysis. *Proc. IEEE*, 57(7):1408–1418, August 1969.

17. I. Claesson and S. Nordholm. A Spatial Filtering Approach to Robust Adaptive Beaming. *IEEE Trans. Antennas Propagat.*, 40(9):1093–1096, September 1992.

18. I. Cohen, S. Gannot, and B. Berdugo. An integrated real-time beamforming and postfiltering system for non-stationary noise environments. *EURASIP Journal on Applied Signal Processing*, (11):1064–1073, October 2003.

19. P. Comon. Independent component analysis, a new concept? *Signal Processing*, 36(3):287–314, April 1994.

20. B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters. Analysis of localization cue preservation by multichannel Wiener filtering based binaural noise reduction in hearing aids. In *Proc. European Signal Processing Conf. (EUSIPCO)*, Lausanne, Switzerland, August 2008.

21. B. Cornelis, S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen. Theoretical analysis of binaural multi-microphone noise reduction techniques. *submitted to IEEE Trans. Audio, Speech and Language Processing*, 2008.

22. H. Cox. Resolving power and sensitivity to mismatch of optimum array processors. *Journal of the Acoustical Society of America*, 54(3):771–785, September 1973.

23. H. Cox, R. M. Zeskind, and M. M. Owen. Robust Adaptive Beamforming. *IEEE Trans. Acoust., Speech, Signal Processing*, 35(10):1365–1376, October 1987.

24. R. E. Crochiere. A weighted overlap-add method for short-time fourier analysis/synthesis. *IEEE Trans. Acoust., Speech, Signal Processing*, 28(1):99–102, February 1980.

25. B. de Vries and R. A. J. de Vries. An integrated approach to hearing aid algorithm design for enhancement of audibility, intelligibility and comfort. In *Proc. of the IEEE Benelux Signal Processing Symposium (SPS2004)*, pages 65–68, Hilvarenbeek, The Netherlands, April 2004.

26. J.G. Desloge, W.M. Rabinowitz, and P.M. Zurek. Microphone-array hearing aids with binaural output–Part I: Fixed-processing systems. *IEEE Trans. Speech and Audio Processing*, 5(6):529–542, November 1997.

27. S. Doclo, R. Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen. Extension of the multi-channel Wiener filter with localisation cues for noise reduction in binaural hearing aids. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pages 221–224, Eindhoven, The Netherlands, September 2005.

28. S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen. Theoretical analysis of binaural cue preservation using multi-channel Wiener filtering and interaural transfer functions. In *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, France, September 2006.

29. S. Doclo and M. Moonen. GSVD-based optimal filtering for single and multimicrophone speech enhancement. *IEEE Trans. Signal Processing*, 50(9):2230–2244, September 2002.

30. S. Doclo and M. Moonen. Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics. *IEEE Trans. Signal Processing*, 51(10):2511–2526, October 2003.

31. S. Doclo and M. Moonen. Design of far-field and near-field broadband beamformers using eigenfilters. *Signal Processing*, 83(12):2641–2673, December 2003.

32. S. Doclo and M. Moonen. On the Output SNR of the Speech-Distortion Weighted Multichannel Wiener Filter. *IEEE Signal Processing Lett.*, 12(12):809–811, December 2005.

33. S. Doclo, A. Spriet, J. Wouters, and M. Moonen. Frequency-Domain Criterion for Speech Distortion Weighted Multichannel Wiener Filter for Robust Noise Reduction. *Speech Communication*, 49(7–8):636–656, Jul.-Aug. 2007.

34. S. Doclo, T. Van den Bogaert, J. Wouters, and M. Moonen. Comparison of reduced-bandwidth MWF-based noise reduction algorithms for binaural hearing aids. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 223–226, New Paltz NY, USA, October 2007.

35. G. Elko. *Superdirectional Microphone Arrays*, chapter 10 in "Acoustic Signal Processing for Telecommunication" (Gay, S. L. and Benesty, J., Eds.), pages 181–237. Kluwer Academic Publishers, Boston, 2000.

36. D. Ellis. *Prediction-driven computational auditory scene analysis*. PhD thesis, MIT, USA, 1996.

37. Y. Ephraim and D. Malah. Speech Enhancement Using a Minimun Mean-Square Error Short-Time Spectral Amplitude Estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 32(6):1109–1121, December 1984.

38. Y. Ephraim and D. Malah. Speech Enhancement Using a Minimun Mean-Square Error Log-Spectral Amplitude Estimator. *IEEE Trans. Acoust., Speech, Signal Processing*, 33(2):443–445, April 1985.

39. O. L. Frost III. An Algorithm for Linearly Constrained Adaptive Array Processing. *Proc. IEEE*, 60(8):926–935, August 1972.

40. S. Gannot, D. Burshtein, and E. Weinstein. Signal Enhancement Using Beamforming and Non-Stationarity with Applications to Speech. *IEEE Trans. Signal Processing*, 49(8):1614–1626, August 2001.

41. S. Gannot, D. Burshtein, and E. Weinstein. Analysis of the Power Spectral Deviation of the General Transfer Function GSC. *IEEE Trans. Signal Processing*, 52(4):1115–1121, April 2004.

42. S. Gannot and I. Cohen. *Springer Handbook of Speech Processing*, chapter Adaptive Beamforming and Postfitering. Springer, November 2007.

43. N. Grbíc and S. Nordholm. Soft contrained subband beamforming for hands-free speech enhancement. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 885–888, Orlando FL, USA, May 2002.

44. J. E. Greenberg, P. M. Peterson, and P. M. Zurek. Intelligibility-weighted measures of speech-to-interference ratio and speech system performance. *Journal of the Acoustical Society of America*, 94(5):3009–3010, November 1993.

45. J. E. Greenberg and P. M. Zurek. Evaluation of an adaptive beamforming method for hearing aids. *Journal of the Acoustical Society of America*, 91(3):1662–1676, March 1992.

46. L. J. Griffiths and C. W. Jim. An alternative approach to linearly constrained adaptive beamforming. *IEEE Trans. Antennas Propagat.*, 30(1):27–34, January 1982.

47. V. Hamacher, J. Chalupper, J. Eggers, E. Fischer, U. Kornagel, H. Puder, and U. Rass. Signal processing in high-end hearing aids: State of the art, challenges, and future trends. *EURASIP Journal on Applied Signal Processing*, 2005(18):2915–2929, 2005.

48. M. L. Hawley, R. Y. Litovsky, and Colburn. Speech intelligibility and localization in a multisource environment. *Journal of the Acoustical Society of America*, 105(6):3436–3448, June 1999.

49. S. Haykin. *Adaptive Filter Theory*. Information and system sciences series. Prentice Hall, Englewood Cliffs, New Jersey, 3rd edition, 1996.

50. W. Herbordt and W. Kellermann. Frequency-domain integration of acoustic echo cancellation and a Generalized Sidelobe Canceller with improved robustness. *European Transactions on Telecommunications*, 13(2):123–132, Mar.-Apr. 2002.

51. W. Herbordt and W. Kellermann. *Adaptive Beamforming for Audio Signal Acquisition*, chapter 6 in "Adaptive Signal Processing: Applications to Real-World Problems" (Benesty, J. and Huang, Y. , Eds.), pages 155–194. Springer-Verlag, 2003.

52. M. W. Hoffman and K. M. Buckley. Robust time-domain processing of broadband microphone array data. *IEEE Trans. Speech and Audio Processing*, 3(3):193–203, May 1995.

53. O. Hoshuyama, B. Begasse, and A. Sugiyama. A new adaptation-mode control based on cross correlation for a robust adaptive microphone array. *IEICE Trans. Fundamentals*, E84-A(2):406–413, February 2001.

54. O. Hoshuyama, A. Sugiyama, and A. Hirano. A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Trans. Signal Processing*, 47(10):2677–2684, October 1999.

55. G. N. Hu and D. L. Wang. Monaural speech segregation based on pitch tracking and amplitude modulation. *IEEE Trans. on Neural Networks*, 15(5):1135–1150, September 2004.

56. N. K. Jablon. Adaptive beamforming with the Generalized Sidelobe Canceller in the presence of array imperfections. *IEEE Trans. Antennas Propagat.*, 34(8):996–1012, August 1986.

57. E. Jan and J. Flanagan. Microphone arrays for speech processing. In *Proc. URSI International Symposium on Signals, Systems, and Electronics*, pages 373–376, San Francisco CA, USA, October 1995.

58. 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. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 917–920, Atlanta GA, USA, May 1996.

59. M. Kajala and M. Hämäläinen. Broadband beamforming optimization for speech enhancement in noisy environments. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pages 19–22, New Paltz NY, USA, October 1999.

60. J. M. Kates. Superdirective arrays for hearing aids. *Journal of the Acoustical Society of America*, 94(4):1930–1933, October 1993.

61. J. M. Kates and M. R. Weiss. A comparison of hearing-aid array-processing techniques. *Journal of the Acoustical Society of America*, 99(5):3138–3148, May 1996.

62. T.J. Klasen, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters. Binaural multi-channel Wiener filtering for hearing aids: preserving interaural time and level differences. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 145–148, Toulouse, France, May 2006.

63. T.J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters. Binaural noise reduction algorithms for hearing aids that preserve interaural time delay cues. *IEEE Trans. Signal Processing*, 55(4):1579–1585, April 2007.

64. B. Kollmeier and R. Koch. Speech enhancement based on physiological and psychoacoustical modals of modulation perception and binaural interaction. *Journal of the Acoustical Society of America*, 95(3):1593–1602, March 1994.

65. M. Kompis and N. Dillier. Performance of an adaptive beamforming noise reduction scheme for hearing aid applications. *Journal of the Acoustical Society of America*, 109(3):1123–1143, 2001.

66. T. Lotter. *Single and multimicrophone speech enhancement for hearing aids.* PhD thesis, RWTH Aachen, Germany, August 2004.

67. F. L. Luo, J. Y. Yang, C. Pavlovic, and A. Nehorai. Adaptive null-forming scheme in digital hearing aids. *IEEE Trans. Signal Processing*, 50(7):1583–1590, 2002.

68. J.-B. Maj, M. Moonen, and J. Wouters. SVD-based optimal filtering technique for noise reduction in hearing aids using two microphones. *EURASIP Journal on Applied Signal Processing*, 2002(4):432–443, April 2002.

69. J.-B. Maj, J. Wouters, and M. Moonen. Noise reduction results of an adaptive filtering technique for dual-microphone behind-the-ear hearing aids. *Ear and Hearing*, 25(3):215–229, June 2004.

70. S. Makino. *Adaptive Signal Processing: Applications to Real-World Problems*, chapter Blind source separation of convolutive mixtures of speech. Springer-Verlag, 2003.

71. C. Marro, Y. Mahieux, and K. U. Summer. Analysis of Noise Reduction and Dereverberation Techniques Based on Microphone Arrays with Postfiltering. *IEEE Trans. Speech and Audio Processing*, 6(3):240–259, May 1998.

72. I.L.D.M. Merks, M.M. Boone, and A.J. Berkhout. Design of a broadside array for a binaural hearing aid. In *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, New Paltz NY, USA, October 1997.

73. T. Nakatani and H. G. Okuno. Harmonic sound stream segregation using localisation and its application to speech stream segregation. *Speech Communication*, 27(3–4):209–222, April 1999.

74. M. Nilsson, S. D. Soli, and A. Sullivan. Development of the Hearing in Noise Test for the measurement of speech reception thresholds in quiet and in noise. *Journal of the Acoustical Society of America*, 95(2):1085–1099, February 1994.

75. R. Nishimura, Y. Suzuki, and F. Asano. A new adaptive binaural microphone array system using a weighted least squares algorithm. In *Proc. IEEE Int. Conf. Acoustics,*

Speech, and Signal Processing (ICASSP), pages 1925–1928, Orlando FL, USA, May 2002.

76. S. Nordebo, I. Claesson, and S. Nordholm. Adaptive beamforming: Spatial filter designed blocking matrix. *IEEE Journal of Oceanic Engineering*, 19(4):583–590, October 1994.
77. S. Nordholm, I. Claesson, and M. Dahl. Adaptive microphone array employing calibration signals: an analytical evaluation. *IEEE Trans. Speech and Audio Processing*, 7(3):241–252, May 1999.
78. S. Nordholm, I. Claesson, and P. Eriksson. The broadband Wiener solution for Griffiths-Jim beamformers. *IEEE Trans. Signal Processing*, 40(2):474–478, February 1992.
79. S. Nordholm, I. Claesson, and N. Grbíc. *Optimal and adaptive microphone arrays for speech input in automobiles*, chapter 14 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 307–326. Springer-Verlag, May 2001.
80. Acoustical Society of America. ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index, June 1997.
81. L. Parra and C. Spence. Convolutive Blind Separation of Non-Stationary Sources. *IEEE Trans. Speech and Audio Processing*, 8(3):320–327, May 2000.
82. D. Rabinkin, R. Renomeron, J. Flanagan, and D. F. Macomber. Optimal truncation time for matched filter array processing. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 3269–3272, Seattle WA, USA, May 1998.
83. G. Reuven, S. Gannot, and I. Cohen. Dual-source transfer-function generalized sidelobe canceller. *IEEE Trans. Audio, Speech and Language Processing*, 16(4):711–727, May 2008.
84. N. Roman, D. L. Wang, and G. J. Brown. Speech segregation based on sound localization. *Journal of the Acoustical Society of America*, 114(4):2236–2252, October 2003.
85. G. Rombouts and M. Moonen. QRD-based unconstrained optimal filtering for acoustic noise reduction. *Signal Processing*, 83(9):1889–1904, September 2003.
86. D. F. Rosenthal and H. G. Okun. *Computational Auditory Scene Analysis*. Lawrence Erlbaum Associates, 1998.
87. O. Shalvi and E. Weinstein. System identification using nonstationary signals. *IEEE Trans. Signal Processing*, 44(8):2055–2063, August 1996.
88. J. J. Shynk. Frequency-domain and multirate and adaptive filtering. *IEEE Signal Processing Magazine*, 9(1):14–37, January 1992.
89. K. U. Simmer, J. Bitzer, and C. Marro. *Post-filtering techniques*, chapter 3 in "Microphone Arrays: Signal Processing Techniques and Applications" (Brandstein, M. S. and Ward, D. B., Eds.), pages 39–57. Springer-Verlag, May 2001.
90. W. Soede, A. J. Berkhout, and F. A. Bilsen. Development of a directional hearing instrument based on array technology. *Journal of the Acoustical Society of America*, 94(2):785–798, August 1993.
91. A. Spriet, M. Moonen, and J. Wouters. Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction. *Signal Processing*, 84(12): 2367–2387, December 2004.
92. A. Spriet, M. Moonen, and J. Wouters. The impact of speech detection errors on the noise reduction performance of multi-channel Wiener filtering and Generalized Sidelobe Cancellation. *Signal Processing*, 85(6):1073–1088, June 2005.

93. A. Spriet, M. Moonen, and J. Wouters. Robustness Analysis of Multi-channel Wiener Filtering and Generalized Sidelobe Cancellation for Multi-microphone Noise Reduction in Hearing Aid Applications. *IEEE Trans. Speech and Audio Processing*, 13(4):487–503, July 2005.

94. A. Spriet, M. Moonen, and J. Wouters. Stochastic gradient-based implementation of spatially preprocessed speech distortion weighted multichannel Wiener filtering for noise reduction in hearing aids. *IEEE Trans. Signal Processing*, 53(3):911–925, March 2005.

95. Z. Tian, K.L. Bell, and H.L. Van Trees. A recursive least squares implementation for LCMP beamforming under quadratic constraint. *IEEE Trans. Signal Processing*, 49(6):1138–1145, June 2001.

96. D. Van Compernolle. Switching Adaptive Filters for Enhancing Noisy and Reverberant Speech from Microphone Array Recordings. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 833–836, Albuquerque NM, USA, April 1990.

97. T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters. Binaural cue preservation for hearing aids using an interaural transfer function multichannel Wiener filter. In *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pages 565–568, Honolulu HI, USA, April 2007.

98. T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters. The effect of multimicrophone noise reduction systems on sound source localization by users of binaural hearing aids. *Journal of the Acoustical Society of America*, 124(1):484–497, July 2008.

99. T. Van den Bogaert, T. J. Klasen, M. Moonen, L. Van Deun, and J. Wouters. Horizontal localisation with bilateral hearing aids: without is better than with. *Journal of the Acoustical Society of America*, 119(1):515–526, January 2006.

100. S. Van Gerven and F. Xie. A comparative study of speech detection methods. In *Proc. EUROSPEECH*, volume 3, pages 1095–1098, Rhodos, Greece, 1997.

101. B. D. Van Veen and K. M. Buckley. Beamforming: A Versatile Approach to Spatial Filtering. *IEEE ASSP Magazine*, pages 4–24, April 1988.

102. J. Vanden Berghe and J. Wouters. An adaptive noise canceller for hearing aids using two nearby microphones. *Journal of the Acoustical Society of America*, 103(6):3621–3626, June 1998.

103. D. B. Ward, R. A. Kennedy, and R. C. Williamson. Theory and design of broadband sensor arrays with frequency invariant far-field beam patterns. *Journal of the Acoustical Society of America*, 97(2):91–95, February 1995.

104. D.P. Welker, J.E. Greenberg, J.G Desloge, and P.M. Zurek. Microphone-array hearing aids with binaural output–Part II: A two-microphone adaptive system. *IEEE Trans. Speech and Audio Processing*, 5(6):543–551, November 1997.

105. T. Wittkop. *Two-channel noise reduction algorithms motivated by models of binaural interaction*. PhD thesis, Universität Oldenburg, Germany, March 2001.