1 Speech Distortion Weighted Multichannel Wiener Filtering Techniques for Noise Reduction

Simon Doclo[1], Ann Spriet[1,2], Jan Wouters[2], and Marc Moonen[1]

[1] Katholieke Universiteit Leuven, Dept. of Electrical Engineering (ESAT-SCD)
Leuven 3001, Belgium
E-mail: {doclo, spriet, moonen}@esat.kuleuven.ac.be
[2] Katholieke Universiteit Leuven, Laboratory for Exp. ORL
Leuven 3000, Belgium
E-mail: jan.wouters@uz.kuleuven.ac.be

Abstract. In many speech communication applications a microphone array is available nowadays, such that multi-microphone speech enhancement techniques can be used instead of single-microphone speech enhancement techniques. A well-known multi-microphone speech enhancement technique is the generalized sidelobe canceller (GSC), which is however quite sensitive to signal model errors, such as microphone mismatch. This chapter discusses a more robust technique called the spatially pre-processed speech distortion weighted multichannel Wiener filter (SP-SDW-MWF), which takes speech distortion due to signal model errors explicitly into account in its design criterion, and which encompasses the standard GSC as a special case. In addition, a novel frequency-domain criterion for the SDW-MWF is presented, from which several – existing and novel – adaptive frequency-domain algorithms can be derived for implementing the SDW-MWF. The noise reduction performance and the robustness of these adaptive algorithms is investigated for a hearing aid application. Using experimental results with a small-sized microphone array, it is shown that the SP-SDW-MWF is more robust against signal model errors than the GSC, both in stationary and in changing noise scenarios.

1.1 Introduction

In many speech communication applications, such as hands-free mobile telephony, hearing aids and voice-controlled systems, the recorded speech signals are often corrupted by a considerable amount of acoustic background noise. Generally speaking, background noise is broadband and non-stationary, and the signal-to-noise ratio (SNR) of the microphone signals may be quite low. Background noise causes a signal degradation that can lead to total unintelligibility of the speech signal and that substantially decreases the performance of speech coding and speech recognition systems. Therefore efficient speech enhancement techniques are required.

Since the desired speech signal and the undesired noise signal usually occupy overlapping frequency bands, single-microphone speech enhancement techniques, such as spectral subtraction, Kalman filtering, and signal

subspace-based techniques, may encounter problems to reduce the background noise without introducing noticeable artifacts (e.g. musical noise) or speech distortion. When the speech and the noise sources are physically located at different positions, it is possible to exploit this spatial diversity by using a microphone array, such that both the spectral and the spatial characteristics can be used in the speech enhancement algorithm.

Well-known multi-microphone speech enhancement techniques are fixed and adaptive beamforming [1]. In a minimum variance distortionless response (MVDR) beamformer [2], the energy of the output signal is minimized under the constraint that signals arriving from the look direction, i.e. the direction of the speech source, are processed without distortion. A widely studied adaptive implementation of this beamformer is the generalized sidelobe canceller (GSC) [3,4], which consists of a fixed spatial pre-processor, i.e. a fixed beamformer, creating a so-called speech reference, and a blocking matrix, creating so-called noise references; and a multichannel adaptive noise canceller, eliminating the noise components in the speech reference which are correlated with the noise references. Multi-microphone noise reduction techniques based on the GSC have already been successfully implemented for various speech applications [5–10].

Due to room reverberation, microphone mismatch, look direction error and spatially distributed sources, undesired speech components however leak into the noise references of the standard GSC, giving rise to speech distortion and cancellation. Several techniques have already been proposed for limiting the speech distortion that results from this speech leakage, by

- reducing the speech leakage components in the noise references, e.g. by using a robust fixed spatial filter designed blocking matrix [5,11–13], by using an adaptive blocking matrix [14–17], or by constructing the blocking matrix based on estimating the acoustic transfer functions from the speech source to the microphone array [18,19];
- limiting the distorting effect of the present speech leakage by
  - updating the adaptive filter only during periods (and for frequencies) where the noise component is dominant, i.e. low SNR [5–7,10,14–17,20]; and
  - constraining the update formula of the adaptive filter, e.g. by imposing a quadratic inequality constraint (QIC) [21–24], by using the leaky least mean square (LMS) algorithm [11,12], by using coefficient constraints [15], or by taking speech distortion due to speech leakage directly into account using *speech distortion weighted multichannel Wiener filtering* (SDW-MWF) [25–27].

In [25], a generalized noise reduction scheme, called the spatially preprocessed speech distortion weighted multichannel Wiener filter (SP-SDW-MWF), has been presented, which consists of a fixed spatial pre-processor and an adaptive SDW-MWF stage. By taking speech distortion explicitly into account in the design criterion of the adaptive stage, the SP-SDW-MWF

adds robustness to the GSC. In [25], it has been shown that, compared to the widely studied QIC-GSC, the SP-SDW-MWF achieves a better noise reduction performance for a given maximum speech distortion level.

In [28–31], recursive matrix-based implementations for the SDW-MWF have been proposed based on the generalized singular value decomposition or the QR decomposition, which are computationally quite expensive. In [26,32] cheaper (time-domain and frequency-domain) stochastic gradient algorithms have been presented. These algorithms however require large circular data buffers, resulting in a large memory usage. In [27,33] adaptive frequency-domain algorithms for the SDW-MWF have been presented using (block-)diagonal correlation matrices, reducing the memory usage and the computational complexity.

In this chapter, we present a novel *frequency-domain criterion* for the SDW-MWF, trading off noise reduction and speech distortion. This frequency-domain criterion for multichannel speech enhancement is in fact an extension of the criterion used in [34–36] for multichannel acoustic echo cancellation. Using the proposed criterion, several existing [27,33] and novel adaptive frequency-domain algorithms for the SDW-MWF can be derived. The main difference between these algorithms consists in the calculation of the step size matrix in the update formula for the adaptive filter and in the calculation of the regularization term (cf. Sects. 1.3 and 1.4).

The chapter is organized as follows. In Sect. 1.2, the GSC and the spatially pre-processed SDW-MWF are briefly reviewed. In Sect. 1.3, the novel frequency-domain criterion for the SDW-MWF is presented. An adaptive algorithm is derived for optimizing this criterion and it is shown how this adaptive algorithm can be implemented in practice. In Sect. 1.4, several approximations are proposed for reducing the computational complexity, where some of these approximations lead to already existing frequency-domain algorithms for the SDW-MWF [27,33]. In Sect. 1.5, the noise reduction performance, the robustness against signal model errors, and the tracking behaviour of the proposed algorithms are illustrated using experimental results for a small-sized microphone array in a hearing aid.

1.2 GSC and Spatially Pre-Processed SDW-MWF

1.2.1 Notation and General Structure

Consider a microphone array with M microphones, where each microphone signal $u_i[k]$, $i = 1 \dots M$, at time k, consists of a filtered version of the clean speech signal $s[k]$ and additive noise, i.e.

$$u_i[k] = h_i[k] \otimes s[k] + u_i^v[k],\ i = 1 \dots M, \tag{1.1}$$

with $h_i[k]$ the acoustic impulse response between the speech source and the ith microphone and $\otimes$ denoting convolution. The additive noise $u_i^v[k]$ can be colored and is assumed to be uncorrelated with the clean speech signal.

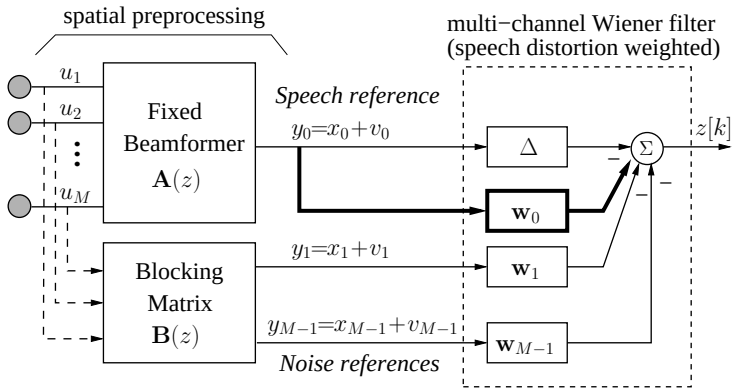

Fig. 1.1. Structure of the spatially pre-processed speech distortion weighted multichannel Wiener filter (SP-SDW-MWF).

The spatially pre-processed speech distortion weighted multichannel Wiener Filter (SP-SDW-MWF) [25] is depicted in Fig. 1.1. It consists of a fixed spatial pre-processor, i.e. a fixed beamformer and a blocking matrix, and an adaptive stage. Note that the structure of the SP-SDW-MWF strongly resembles the standard GSC, where the difference lies in the fact that an adaptive SDW-MWF is used in the adaptive stage and that it is possible to include an extra filter $\mathbf{w}_0$ on the speech reference.

The *fixed beamformer* $\mathbf{A}(z)$ creates a so-called speech reference

$$y_0[k] = x_0[k] + v_0[k], \tag{1.2}$$

with $x_0[k]$ and $v_0[k]$ respectively the speech and the noise component of the speech reference, by steering a beam towards the direction of the speaker. The fixed beamformer should be designed such that the distortion of the speech component $x_0[k]$, due to possible errors in the assumed signal model (e.g. look direction error, microphone mismatch) is small. A delay-and-sum beamformer, which time-aligns the microphone signals, offers sufficient robustness against signal model errors since it minimizes the noise sensitivity. However, in order to achieve a better spatial selectivity while preserving robustness, the fixed beamformer can be optimized, e.g. using statistical knowledge about the signal model errors that occur in practice [13].

The *blocking matrix* $\mathbf{B}(z)$ creates $M-1$ so-called noise references

$$y_n[k] = x_n[k] + v_n[k],\ n = 1 \dots M-1, \tag{1.3}$$

by steering zeroes towards the direction of the speaker, such that it is anticipated that the noise components $v_n[k]$ are dominant compared to the speech components $x_n[k]$. A simple technique to create the noise references consists of pair-wisely subtracting the time-aligned microphone signals. Under ideal conditions (i.e. no reverberation, point speech source, no look direction error,

no microphone mismatch), the noise references only contain noise components $v_n[k]$. Since these conditions are practically never fulfilled, undesired speech components $x_n[k]$, i.e. so-called *speech leakage*, are present in the noise references. Several techniques have already been proposed for reducing the speech leakage components in the noise references [5,11–19].

During *speech-periods* the speech and the noise references consist of speech+noise, i.e. $y_n[k] = x_n[k] + v_n[k]$, whereas during *noise-only-periods* only the noise components $v_n[k]$ are observed. We assume that the (spatial and/or temporal) second-order statistics of the noise are sufficiently stationary such that they can be estimated during noise-only-periods and used during subsequent speech-periods. This requires the use of a voice activity detection (VAD) mechanism [37–39] or an on-line procedure for estimating the SNR [40].

The goal of the *adaptive stage* is to make an estimate of the noise component in the speech reference and to subtract this noise estimate from the speech reference in order to obtain an enhanced output signal $z[k]$. Let N be the number of input channels to the multichannel filter ($N = M$ if the filter $\mathbf{w}_0$ on the speech reference is present, $N = M - 1$ otherwise). Let the FIR filters $\mathbf{w}_n[k]$, $n = M - N \dots M - 1$, have length L, and consider the L-dimensional data vectors $\mathbf{y}_n[k]$, the NL-dimensional stacked data vector $\mathbf{y}[k]$, and the NL-dimensional stacked filter $\mathbf{w}[k]$, defined as

$$\mathbf{y}_n[k] = \big[\, y_n[k]\ y_n[k-1] \dots y_n[k-L+1] \,\big]^T, \quad n = M-N \dots M-1, \tag{1.4}$$

$$\mathbf{y}[k] = \big[\, \mathbf{y}_{M-N}^T[k]\ \mathbf{y}_{M-N+1}^T[k] \cdots \mathbf{y}_{M-1}^T[k] \,\big]^T, \tag{1.5}$$

$$\mathbf{w}[k] = \big[\, \mathbf{w}_{M-N}^T[k]\ \mathbf{w}_{M-N+1}^T[k] \dots \mathbf{w}_{M-1}^T[k] \,\big]^T, \tag{1.6}$$

with ^{T} denoting transpose of a vector or a matrix. The stacked data vector can be decomposed into a speech and a noise component, i.e. $\mathbf{y}[k] = \mathbf{x}[k] + \mathbf{v}[k]$, where $\mathbf{x}[k]$ and $\mathbf{v}[k]$ are defined similarly as in (1.5). The goal of the filter $\mathbf{w}[k]$ is to make an estimate of the delayed noise component $v_0[k - \Delta]$ in the speech reference[1]. This noise estimate is then subtracted from the speech reference in order to obtain the enhanced output signal $z[k]$, i.e.

$$z[k] = y_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{y}[k] \tag{1.7}$$

$$= x_0[k-\Delta] + \underbrace{(v_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{v}[k])}_{e_v[k]} - \underbrace{\mathbf{w}^T[k]\mathbf{x}[k]}_{e_x[k]}. \tag{1.8}$$

Hence, the output signal $z[k]$ consists of 3 terms: the delayed speech component $x_0[k - \Delta]$ in the speech reference, residual noise $e_v[k]$, and (linear) speech distortion $e_x[k]$. The goal of any speech enhancement algorithm is to

[1] The delay Δ is applied to the speech reference in order to allow for non-causal filter taps. This delay is usually set equal to $\lceil L/2 \rceil$, where $\lceil x \rceil$ denotes the smallest integer larger than or equal to x.

reduce the residual noise as much as possible, while simultaneously limiting the speech distortion. The speech distortion can e.g. be limited by reducing the speech leakage components $\mathbf{x}[k]$ in the noise references and/or by constraining the filter $\mathbf{w}[k]$.

1.2.2 Generalized Sidelobe Canceller

The standard generalized sidelobe canceller (GSC) minimizes the residual noise energy without taking into account speech distortion, i.e.

$$J_{GSC}(\mathbf{w}[k]) = \varepsilon_v^2[k] = \mathcal{E}\Big\{\big|v_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{v}[k]\big|^2\Big\}, \tag{1.9}$$

with $\mathcal{E}$ denoting the expected value operator. The filter $\mathbf{w}[k]$ minimizing this cost function is equal to

$$\mathbf{w}[k] = \mathcal{E}\big\{\mathbf{v}[k]\mathbf{v}^T[k]\big\}^{-1}\,\mathcal{E}\big\{\mathbf{v}[k]v_0[k-\Delta]\big\}, \tag{1.10}$$

where the noise correlation matrix $\mathcal{E}\{\mathbf{v}[k]\mathbf{v}^T[k]\}$ and the noise cross-correlation vector $\mathcal{E}\{\mathbf{v}[k]v_0[k-\Delta]\}$ are estimated during noise-only-periods. Hence, in an adaptive implementation, the filter $\mathbf{w}[k]$ is also allowed to be updated only during noise-only-periods [5–7,10,14–17,20], since adaptation during speech-periods would lead to an incorrect solution and signal cancellation. Note however that signal distortion due to speech leakage still occurs even when the adaptive filter is updated only during noise-only-periods, since the speech distortion term $e_x[k]$ is still present in the output signal $z[k]$.

A commonly used approach to increase the robustness against signal model errors is to apply a quadratic inequality constraint (QIC) [21–24], i.e.

$$\mathbf{w}^T[k]\mathbf{w}[k] \leq \beta^2. \tag{1.11}$$

The QIC avoids excessive growth of the filter coefficients $\mathbf{w}[k]$, and hence limits speech distortion $\mathbf{w}^T[k]\mathbf{x}[k]$ due to speech leakage. The QIC can be implemented using the scaled projection algorithm [22] or by using variable loading [24]. Similar approaches for constraining the filter coefficients consist in using the leaky LMS algorithm [11,12] or using coefficient constraints [15].

In the GSC the number of input channels to the adaptive filter is typically equal to $N = M - 1$. It is not possible to include the filter $\mathbf{w}_0$ on the speech reference, since in this case the filter $\mathbf{w}[k]$ in (1.10) would be equal to

$$\mathbf{w}_0[k] = \mathbf{u}_{\Delta+1}, \quad \mathbf{w}_n[k] = \mathbf{0},\ n = 1 \ldots M-1\ , \tag{1.12}$$

with $\mathbf{u}_l$ the lth canonical L-dimensional vector, whose lth element is equal to 1 and all other elements are equal to 0, such that the output signal $z[k] = 0$.

1.2.3 Speech Distortion Weighted Multichannel Wiener Filter

The speech distortion weighted multichannel Wiener filter (SDW-MWF) takes speech distortion due to speech leakage explicitly into account in the design criterion of the filter $\mathbf{w}[k]$ and minimizes the weighted sum of the residual noise energy $\varepsilon_v^2[k]$ and the speech distortion energy $\varepsilon_x^2[k]$, i.e.

$$J(\mathbf{w}[k]) = \varepsilon_v^2[k] + \frac{1}{\mu}\varepsilon_x^2[k] \tag{1.13}$$

$$= \mathcal{E}\Big\{\big|v_0[k-\Delta] - \mathbf{w}^T[k]\mathbf{v}[k]\big|^2\Big\} + \frac{1}{\mu}\mathcal{E}\Big\{\big|\mathbf{w}^T[k]\mathbf{x}[k]\big|^2\Big\}, \tag{1.14}$$

where the parameter $\mu \in [0, \infty]$ provides a trade-off between noise reduction and speech distortion [25,29,41]. If $\mu = 1$, the minimum mean square error (MMSE) criterion is obtained. If $\mu < 1$, speech distortion is reduced at the expense of increased residual noise energy. On the other hand, if $\mu > 1$, residual noise is reduced at the expense of increased speech distortion.

The filter $\mathbf{w}[k]$ minimizing the cost function in (1.14) is equal to

$$\mathbf{w}[k] = \left[\mathcal{E}\big\{\mathbf{v}[k]\mathbf{v}^T[k]\big\} + \frac{1}{\mu}\mathcal{E}\big\{\mathbf{x}[k]\mathbf{x}^T[k]\big\}\right]^{-1} \mathcal{E}\big\{\mathbf{v}[k]v_0[k-\Delta]\big\}, \tag{1.15}$$

where, using the independence assumption between speech and noise, the correlation matrix $\mathcal{E}\{\mathbf{x}[k]\mathbf{x}^T[k]\}$ can be computed as

$$\mathcal{E}\big\{\mathbf{x}[k]\mathbf{x}^T[k]\big\} = \mathcal{E}\big\{\mathbf{y}[k]\mathbf{y}^T[k]\big\} - \mathcal{E}\big\{\mathbf{v}[k]\mathbf{v}^T[k]\big\}. \tag{1.16}$$

The speech correlation matrix $\mathcal{E}\{\mathbf{y}[k]\mathbf{y}^T[k]\}$ is estimated during speech-periods and the noise correlation matrix $\mathcal{E}\{\mathbf{v}[k]\mathbf{v}^T[k]\}$ is estimated during noise-only-periods. We assume that the second-order statistics of the noise are sufficiently stationary such that they can be estimated during noise-only-periods and used during subsequent speech-periods.

Since the SDW-MWF takes speech distortion explicitly into account in its optimization criterion, it is possible to include the filter $\mathbf{w}_0$ on the speech reference. Depending on the setting of the parameter μ and the presence/absence of the filter $\mathbf{w}_0$, different algorithms are obtained:

- Without filter $\mathbf{w}_0$ ($N = M - 1$), we obtain the *speech distortion regularized GSC* (SDR-GSC), where the standard optimization criterion of the GSC in (1.9) is supplemented with a regularization term $1/\mu\,\varepsilon_x^2$. For $\mu = \infty$, speech distortion is completely ignored, which corresponds to the standard GSC. For $\mu = 0$, all emphasis is put on speech distortion, such that $\mathbf{w}[k] = \mathbf{0}$ and the output signal $z[k]$ is equal to the delayed speech reference $y_0[k-\Delta]$. Compared to the QIC-GSC, the SDR-GSC is less conservative, since the regularization term in the SDR-GSC is proportional to the actual amount of speech leakage present in the noise references. On the other hand, the constraint value β^2 in the QIC-GSC, cf. (1.11), needs to be chosen based on the largest signal model errors that may occur,

such that the noise reduction performance is compromised even when no or a small amount of speech leakage is present. In the absence of speech leakage, the regularization term in the SDR-GSC equals 0, such that the GSC solution is obtained and hence the noise reduction performance is not compromised. In [25] it has been shown that in comparison with the QIC-GSC, the SDR-GSC obtains a better noise reduction for small model errors, while guaranteeing robustness against large model errors.

- With filter $\mathbf{w}_0$ ($N = M$), we obtain the *spatially pre-processed speech distortion weighted multichannel Wiener filter* (SP-SDW-MWF). For $\mu = 1$, the output signal $z[k]$ is the MMSE estimate of the delayed speech component $x_0[k - \Delta]$ in the speech reference. In [25] it has been shown that, for infinite filter lengths, the performance of the SP-SDW-MWF is not affected by microphone mismatch as long as the speech component in the speech reference remains unaltered by the microphone mismatch. Hence, the extra filter on the speech reference further improves the performance.

In [28–31], recursive matrix-based implementations for the SDW-MWF have been proposed based on the generalized singular value decomposition or the QR decomposition, which are computationally quite expensive. Starting from the cost function in (1.14), a cheaper time-domain stochastic gradient algorithm has been derived in [26][2]. In order to speed up convergence and reduce the computational complexity, this algorithm has been implemented in the frequency-domain. It has been shown in [26] that for highly non-stationary noise, this stochastic gradient algorithm suffers from a large excess error, which can be reduced by low-pass filtering the regularization term, i.e. the part of the gradient estimate that limits speech distortion. The computation of this regularization term however requires the storage of circular data buffers for the speech-and-noise samples and for the noise-only samples. Using these circular data buffers, the filter coefficients $\mathbf{w}[k]$ can be updated both during speech-periods and during noise-only-periods, but the storage of the data buffers unfortunately gives rise to a large memory usage. In [27], the regularization term has been approximated in the frequency-domain, using (diagonal) speech and noise correlation matrices in the frequency-domain. This approximation leads to a drastic decrease in memory usage, while also further reducing the computational complexity but not compromising the noise reduction performance and the robustness against signal model errors.

In the following section, a novel *frequency-domain criterion* for the SDW-MWF is proposed, similar to the cost function in (1.14). This frequency-domain criterion is an extension of the criterion used in [34–36] for multichannel echo cancellation to the problem of multichannel speech enhancement. Furthermore, it provides a way for linking existing adaptive frequency-

[2] In [32] a similar time-domain stochastic gradient algorithm has been presented, which however invokes some independence assumptions that result in a significant performance degradation compared to the algorithm in [26].

domain algorithms for the SDW-MWF [27,33] and for deriving novel adaptive algorithms, as will be shown in Sect. 1.4.

1.3 Frequency-Domain Criterion for SDW-MWF

1.3.1 Frequency-Domain Notation

We define the L-dimensional block signals $\mathbf{e}_v[m]$ and $\mathbf{e}_x[m]$ as

$$\mathbf{e}_v[m] = \big[\, e_v[mL]\; e_v[mL+1] \ldots e_v[mL+L-1] \,\big]^T, \tag{1.17}$$

$$\mathbf{e}_x[m] = \big[\, e_x[mL]\; e_x[mL+1] \ldots e_x[mL+L-1] \,\big]^T, \tag{1.18}$$

with m the block time index. The block signal $\mathbf{e}_v[m]$, representing the residual noise, can be written as

$$\mathbf{e}_v[m] = \mathbf{d}[m] - \sum_{n=M-N}^{M-1} \mathbf{V}_n^T[m]\, \mathbf{w}_n, \tag{1.19}$$

with the L-dimensional block signal $\mathbf{d}[m]$, and the $L \times L$-dimensional Toeplitz matrices $\mathbf{V}_n[m]$, $n = M-N \ldots M-1$, equal to

$$\mathbf{d}[m] = \big[\, v_0[mL-\Delta]\; v_0[mL-\Delta+1] \ldots v_0[mL-\Delta+L-1] \,\big]^T, \tag{1.20}$$

$$\mathbf{V}_n[m] = \big[\, \mathbf{v}_n[mL]\; \mathbf{v}_n[mL+1] \ldots \mathbf{v}_n[mL+L-1] \,\big]. \tag{1.21}$$

It is well known that the filtering operation $\mathbf{V}_n^T[m]\, \mathbf{w}_n$ can be calculated in the frequency domain as [35,42]

$$\mathbf{V}_n^T[m]\, \mathbf{w}_n = \big[\, \mathbf{0}_L\; \mathbf{I}_L \,\big]\, \mathbf{F}_{2L}^{-1}\, \mathbf{D}_{v,n}[m]\, \mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_L \end{bmatrix} \mathbf{w}_n, \tag{1.22}$$

where $\mathbf{0}_L$ represents the $L \times L$-dimensional zero matrix, $\mathbf{I}_L$ represents the $L \times L$-dimensional unity matrix, $\mathbf{F}_{2L}$ is the $2L \times 2L$-dimensional Fourier matrix and $\mathbf{D}_{v,n}[m]$ is a $2L \times 2L$-dimensional diagonal matrix whose elements are the discrete Fourier transform of the first column of

$$\begin{bmatrix} \mathbf{V}_n^T[m-1] \\ \mathbf{V}_n^T[m] \end{bmatrix}, \tag{1.23}$$

i.e. the $2L$-dimensional vector

$$\big[\, v_n[mL-L] \ldots v_n[mL-1]\; v_n[mL] \ldots v_n[mL+L-1] \,\big]^T. \tag{1.24}$$

Hence, combining (1.19) and (1.22), the block signal $\mathbf{e}_v[m]$ can be written as

$$\mathbf{e}_v[m] = \mathbf{d}[m] - \big[\, \mathbf{0}_L\; \mathbf{I}_L \,\big]\, \mathbf{F}_{2L}^{-1} \sum_{n=M-N}^{M-1} \mathbf{D}_{v,n}[m]\, \mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_L \end{bmatrix} \mathbf{w}_n \tag{1.25}$$

$$= \mathbf{d}[m] - \big[\, \mathbf{0}_L\; \mathbf{I}_L \,\big]\, \mathbf{F}_{2L}^{-1}\, \mathbf{U}_v[m]\, \mathbf{w}, \tag{1.26}$$

with the $2L \times NL$-dimensional matrix $\mathbf{U}_v[m]$ defined as

$$\mathbf{U}_v[m] = \left[\mathbf{D}_{v,M-N}[m]\, \mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_L \end{bmatrix} \dots \mathbf{D}_{v,M-1}[m]\, \mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_L \end{bmatrix} \right] \tag{1.27}$$

$$= \mathbf{D}_v[m]\, \mathbf{F}^{10}_{2NL\times NL}, \tag{1.28}$$

and the $2L\times 2NL$-dimensional matrix $\mathbf{D}_v[m]$ and the $2NL\times NL$-dimensional diagonal block-matrix $\mathbf{F}^{10}_{2NL\times NL}$ equal to

$$\mathbf{D}_v[m] = \left[\mathbf{D}_{v,M-N}[m] \dots \mathbf{D}_{v,M-1}[m] \right] \tag{1.29}$$

$$\mathbf{F}^{10}_{2NL\times NL} = \text{diag} \left[\mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_L \end{bmatrix} \dots \mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L \\ \mathbf{0}_L \end{bmatrix} \right]. \tag{1.30}$$

Similarly, the block signal $\mathbf{e}_x[m]$, representing the speech distortion, can be written as

$$\mathbf{e}_x[m] = \left[\mathbf{0}_L \ \mathbf{I}_L \right] \mathbf{F}^{-1}_{2L}\, \mathbf{U}_x[m]\, \mathbf{w} = \left[\mathbf{0}_L \ \mathbf{I}_L \right] \mathbf{F}^{-1}_{2L}\, \mathbf{D}_x[m]\, \mathbf{F}^{10}_{2NL\times NL}\, \mathbf{w}, \tag{1.31}$$

with $\mathbf{U}_x[m]$ and $\mathbf{D}_x[m]$ defined similarly as $\mathbf{U}_v[m]$ and $\mathbf{D}_v[m]$ for the speech component instead as for the noise component.

If we multiply the block signals in (1.26) and (1.31) with the $L \times L$-dimensional Fourier matrix $\mathbf{F}_L$, we obtain the error signals in the frequency domain (denoted by underbars), i.e.

$$\underline{\mathbf{e}}_v[m] = \mathbf{F}_L\, \mathbf{e}_v[m] = \underline{\mathbf{d}}[m] - \mathbf{G}^{01}_{L\times 2L} \mathbf{U}_v[m]\, \mathbf{w}, \tag{1.32}$$

$$\underline{\mathbf{e}}_x[m] = \mathbf{F}_L\, \mathbf{e}_x[m] = \mathbf{G}^{01}_{L\times 2L} \mathbf{U}_x[m]\, \mathbf{w}, \tag{1.33}$$

with $\underline{\mathbf{d}}[m] = \mathbf{F}_L\, \mathbf{d}[m]$ and

$$\mathbf{G}^{01}_{L\times 2L} = \mathbf{F}_L \left[\mathbf{0}_L \ \mathbf{I}_L \right] \mathbf{F}^{-1}_{2L}\,. \tag{1.34}$$

We now define a *frequency-domain criterion* similar to (1.14), minimizing the weighted sum of residual noise energy and speech distortion energy, as

$$J_f[m] = (1-\lambda_v) \sum_{i=0}^{m} \lambda_v^{m-i} \underline{\mathbf{e}}_v^H[i]\, \underline{\mathbf{e}}_v[i] + \frac{1}{\mu}(1-\lambda_x) \sum_{i=0}^{m} \lambda_x^{m-i} \underline{\mathbf{e}}_x^H[i]\, \underline{\mathbf{e}}_x[i], \tag{1.35}$$

where ^{H} denotes complex conjugate of a vector or a matrix, λ_v and λ_x are exponential forgetting factors respectively for noise and speech ($0 < \lambda_v < 1$, $0 < \lambda_x < 1$), and $1/\mu$ is the trade-off parameter between noise reduction and speech distortion.

1.3.2 Normal Equations

The cost function $J_f[m]$ can be minimized by setting its derivative with respect to the (time-domain) filter coefficients $\mathbf{w}[m]$ equal to zero. Using

(1.32) and (1.33), the derivative is equal to

$$\frac{\partial J_f[m]}{\partial \mathbf{w}[m]} = (1-\lambda_v)\sum_{i=0}^{m}\lambda_v^{m-i}\Big(\mathbf{U}_v^H[i]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_v[i]\mathbf{w}[m] - \mathbf{U}_v^H[i]\,\underline{\mathbf{d}}_{2L}[i]\Big) + \frac{1}{\mu}(1-\lambda_x)\sum_{i=0}^{m}\lambda_x^{m-i}\,\mathbf{U}_x^H[i]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_x[i]\mathbf{w}[m], \tag{1.36}$$

with

$$\underline{\mathbf{d}}_{2L}[m] = 2(\mathbf{G}_{L\times 2L}^{01})^H\underline{\mathbf{d}}[m] = \mathbf{F}_{2L}\begin{bmatrix}\mathbf{0}_L\\ \mathbf{I}_L\end{bmatrix}\mathbf{d}[m] \tag{1.37}$$

$$\mathbf{G}_{2L\times 2L}^{01} = 2(\mathbf{G}_{L\times 2L}^{01})^H\mathbf{G}_{L\times 2L}^{01}, = \mathbf{F}_{2L}\begin{bmatrix}\mathbf{0}_L & \mathbf{0}_L\\ \mathbf{0}_L & \mathbf{I}_L\end{bmatrix}\mathbf{F}_{2L}^{-1}. \tag{1.38}$$

Hence, the *normal equations* can be written as

$$\left[\mathbf{S}_v[m] + \frac{1}{\mu}\mathbf{S}_x[m]\right]\mathbf{w}[m] = \mathbf{s}[m], \tag{1.39}$$

with the $NL \times NL$-dimensional correlation matrices $\mathbf{S}_v[m]$ and $\mathbf{S}_x[m]$, and the NL-dimensional cross-correlation vector $\mathbf{s}[m]$ defined as

$$\mathbf{S}_v[m] = (1-\lambda_v)\sum_{i=0}^{m}\lambda_v^{m-i}\,\mathbf{U}_v^H[i]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_v[i] \tag{1.40}$$

$$= \lambda_v\mathbf{S}_v[m-1] + (1-\lambda_v)\,\mathbf{U}_v^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_v[m], \tag{1.41}$$

$$\mathbf{S}_x[m] = (1-\lambda_x)\sum_{i=0}^{m}\lambda_x^{m-i}\,\mathbf{U}_x^H[i]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_x[i] \tag{1.42}$$

$$= \lambda_x\mathbf{S}_x[m-1] + (1-\lambda_x)\,\mathbf{U}_x^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_x[m], \tag{1.43}$$

$$\mathbf{s}[m] = (1-\lambda_v)\sum_{i=0}^{m}\lambda_v^{m-i}\,\mathbf{U}_v^H[i]\,\underline{\mathbf{d}}_{2L}[i] \tag{1.44}$$

$$= \lambda_v\mathbf{s}[m-1] + (1-\lambda_v)\,\mathbf{U}_v^H[m]\,\underline{\mathbf{d}}_{2L}[m]. \tag{1.45}$$

The optimal Wiener filter is the solution of these normal equations.

1.3.3 Adaptive Algorithm

An algorithm for adapting $\mathbf{w}[m]$ can be found by enforcing the normal equations (1.39) at block time m and $m-1$, i.e.

$$\left[\mathbf{S}_v[m] + \frac{1}{\mu}\mathbf{S}_x[m]\right]\mathbf{w}[m] = \lambda_v\left[\mathbf{S}_v[m-1] + \frac{1}{\mu}\mathbf{S}_x[m-1]\right]\mathbf{w}[m-1] + (1-\lambda_v)\,\mathbf{U}_v^H[m]\,\underline{\mathbf{d}}_{2L}[m]$$

$$= \Bigg[\frac{1}{\mu}\frac{\lambda_v}{\lambda_x}\Big(\mathbf{S}_x[m] - (1-\lambda_x)\,\mathbf{U}_x^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_x[m]\Big) + \mathbf{S}_v[m] - (1-\lambda_v)\,\mathbf{U}_v^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_v[m]\Bigg]\mathbf{w}[m-1] + (1-\lambda_v)\,\mathbf{U}_v^H[m]\,\underline{\mathbf{d}}_{2L}[m],$$

such that the update formula for $\mathbf{w}[m]$ can be written as

$$\underline{\mathbf{e}}_{v,2L}[m] = \mathbf{F}_{2L}\begin{bmatrix}\mathbf{0}_L\\ \mathbf{I}_L\end{bmatrix}\mathbf{e}_v[m] = \underline{\mathbf{d}}_{2L}[m] - \mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_v[m]\,\mathbf{w}[m-1], \tag{1.46}$$

$$\underline{\mathbf{e}}_{x,2L}[m] = \mathbf{F}_{2L}\begin{bmatrix}\mathbf{0}_L\\ \mathbf{I}_L\end{bmatrix}\mathbf{e}_x[m] = \mathbf{G}_{2L\times 2L}^{01}\mathbf{U}_x[m]\,\mathbf{w}[m-1], \tag{1.47}$$

$$\mathbf{w}[m] = \left[\mathbf{S}_v[m] + \frac{1}{\mu}\mathbf{S}_x[m]\right]^{-1}\Bigg\{\left[\mathbf{S}_v[m] + \frac{1}{\mu}\frac{\lambda_v}{\lambda_x}\mathbf{S}_x[m]\right]\mathbf{w}[m-1] + (1-\lambda_v)\mathbf{U}_v^H[m]\underline{\mathbf{e}}_{v,2L}[m] - \frac{1}{\mu}\frac{\lambda_v}{\lambda_x}(1-\lambda_x)\mathbf{U}_x^H[m]\underline{\mathbf{e}}_{x,2L}[m]\Bigg\}. \tag{1.48}$$

For convenience, we now define the $2NL\times 2NL$-dimensional correlation matrices $\mathbf{Q}_v[m]$ and $\mathbf{Q}_x[m]$ as

$$\mathbf{S}_v[m] = (\mathbf{F}_{2NL\times NL}^{10})^H\,\mathbf{Q}_v[m]\,\mathbf{F}_{2NL\times NL}^{10}, \tag{1.49}$$

$$\mathbf{S}_x[m] = (\mathbf{F}_{2NL\times NL}^{10})^H\,\mathbf{Q}_x[m]\,\mathbf{F}_{2NL\times NL}^{10}, \tag{1.50}$$

such that

$$\mathbf{Q}_v[m] = \lambda_v\mathbf{Q}_v[m-1] + (1-\lambda_v)\,\mathbf{D}_v^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_v[m], \tag{1.51}$$

$$\mathbf{Q}_x[m] = \lambda_x\mathbf{Q}_x[m-1] + (1-\lambda_x)\,\mathbf{D}_x^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_x[m]. \tag{1.52}$$

In addition, we define the $2NL$-dimensional frequency-domain filter $\underline{\mathbf{w}}_{2NL}[m]$ as

$$\underline{\mathbf{w}}_{2NL}[m] = \mathbf{F}_{2NL\times NL}^{10}\mathbf{w}[m] \tag{1.53}$$

$$= \left[\underline{\mathbf{w}}_{M-N,2L}^T[m]\ \dots\ \underline{\mathbf{w}}_{M-1,2L}^T[m]\right]^T, \tag{1.54}$$

with

$$\underline{\mathbf{w}}_{n,2L}[m] = \mathbf{F}_{2L}\begin{bmatrix}\mathbf{I}_L\\ \mathbf{0}_L\end{bmatrix}\mathbf{w}_n[m]. \tag{1.55}$$

By pre-multiplying both sides of (1.48) with $\mathbf{F}_{2NL\times NL}^{10}$, and by using (1.49) and (1.50), we obtain

$$\underline{\mathbf{e}}_{v,2L}[m] = \underline{\mathbf{d}}_{2L}[m] - \mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_v[m]\,\underline{\mathbf{w}}_{2NL}[m-1], \tag{1.56}$$

$$\underline{\mathbf{e}}_{x,2L}[m] = \mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_x[m]\,\underline{\mathbf{w}}_{2NL}[m-1], \tag{1.57}$$

$$\underline{\mathbf{w}}_{2NL}[m] = \mathbf{F}_{2NL\times NL}^{10}\left[\mathbf{S}_v[m] + \frac{1}{\mu}\mathbf{S}_x[m]\right]^{-1}(\mathbf{F}_{2NL\times NL}^{10})^H\,.$$

$$\left\{\left[\mathbf{Q}_v[m]+\frac{1}{\mu}\frac{\lambda_v}{\lambda_x}\mathbf{Q}_x[m]\right]\underline{\mathbf{w}}_{2NL}[m-1]+(1-\lambda_v)\mathbf{D}_v^H[m]\underline{\mathbf{e}}_{v,2L}[m]\right.$$
$$\left.-\frac{1}{\mu}\frac{\lambda_v}{\lambda_x}(1-\lambda_x)\,\mathbf{D}_x^H[m]\,\underline{\mathbf{e}}_{x,2L}[m]\right\}. \tag{1.58}$$

In [35], it has been shown that

$$\mathbf{F}_{2NL\times NL}^{10}\,\mathbf{S}_v^{-1}[m]\,(\mathbf{F}_{2NL\times NL}^{10})^H=\mathbf{G}_{2NL\times 2NL}^{10}\,\mathbf{Q}_v^{-1}[m], \tag{1.59}$$

with the $2NL\times 2NL$-dimensional diagonal block-matrix $\mathbf{G}_{2NL\times 2NL}^{10}$ defined as

$$\mathbf{G}_{2NL\times 2NL}^{10}=\mathrm{diag}\left[\,\mathbf{G}_{2L\times 2L}^{10}\,\ldots\,\mathbf{G}_{2L\times 2L}^{10}\,\right], \tag{1.60}$$

with

$$\mathbf{G}_{2L\times 2L}^{10}=\mathbf{F}_{2L}\begin{bmatrix}\mathbf{I}_L & \mathbf{0}_L\\ \mathbf{0}_L & \mathbf{0}_L\end{bmatrix}\mathbf{F}_{2L}^{-1}, \tag{1.61}$$

such that (1.58) can be written as

$$\underline{\mathbf{w}}_{2NL}[m]=\mathbf{G}_{2NL\times 2NL}^{10}\left[\mathbf{Q}_v[m]+\frac{1}{\mu}\mathbf{Q}_x[m]\right]^{-1}$$
$$\left\{\left[\mathbf{Q}_v[m]+\frac{1}{\mu}\frac{\lambda_v}{\lambda_x}\mathbf{Q}_x[m]\right]\underline{\mathbf{w}}_{2NL}[m-1]+(1-\lambda_v)\mathbf{D}_v^H[m]\underline{\mathbf{e}}_{v,2L}[m]\right.$$
$$\left.-\frac{1}{\mu}\frac{\lambda_v}{\lambda_x}(1-\lambda_x)\,\mathbf{D}_x^H[m]\,\underline{\mathbf{e}}_{x,2L}[m]\right\}. \tag{1.62}$$

In the following, we will assume that the exponential forgetting factors $\lambda_x=\lambda_v=\lambda$, such that (1.62) reduces to

$$\underline{\mathbf{w}}_{2NL}[m]=\underline{\mathbf{w}}_{2NL}[m-1]+(1-\lambda)\,\mathbf{G}_{2NL\times 2NL}^{10}\left[\mathbf{Q}_v[m]+\frac{1}{\mu}\mathbf{Q}_x[m]\right]^{-1}\cdot$$
$$\left\{\mathbf{D}_v^H[m]\,\underline{\mathbf{e}}_{v,2L}[m]-\frac{1}{\mu}\mathbf{D}_x^H[m]\,\underline{\mathbf{e}}_{x,2L}[m]\right\}. \tag{1.63}$$

When the trade-off parameter $1/\mu=0$, this algorithm is equal to the multichannel frequency-domain adaptive filtering algorithm derived in [35], applied to the GSC. For $1/\mu>0$, the $2NL$-dimensional additional *regularization term*

$$\underline{\mathbf{r}}_{2NL}[m]=\frac{1}{\mu}\mathbf{D}_x^H[m]\,\underline{\mathbf{e}}_{x,2L}[m] \tag{1.64}$$
$$=\frac{1}{\mu}\mathbf{D}_x^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_x[m]\,\underline{\mathbf{w}}_{2NL}[m-1], \tag{1.65}$$

limits speech distortion due to speech leakage in the noise references.

1.3.4 Practical Implementation

Since the SP-SDW-MWF takes speech leakage explicitly into account, it should in principle be possible to adapt the filter $\mathbf{w}[m]$ both during speech-periods and during noise-only-periods, unlike the standard GSC. This approach has been taken in [26], where a stochastic gradient based implementation of the SP-SDW-MWF has been proposed using 2 circular data buffers, i.e. a speech+noise buffer storing speech+noise vectors $\mathbf{y}[k]$ during speech-periods, and a noise-only buffer storing noise-only vectors $\mathbf{v}[k]$ during noise-only-periods. This implementation using circular data buffers however gives rise to a large memory usage.

If we take a closer look at (1.63), we notice that $\mathbf{D}_v[m]$ and $\underline{\mathbf{e}}_{v,2L}[m]$ can only be computed during noise-only-periods, whereas $\mathbf{D}_x[m]$ and $\underline{\mathbf{e}}_{x,2L}[m]$ can only be computed during speech-periods. Hence, we will take a similar approach as in the standard GSC, i.e. *updating the filter coefficients only during noise-only-periods*. Since during noise-only-periods the (instantaneous) correlation matrix $\mathbf{D}_x^H[m]\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_x[m]$ of the clean speech signal, required in the computation of the regularization term $\underline{\mathbf{r}}_{2NL}[m]$, is not available, we will approximate this term by the (average) correlation matrix $\mathbf{Q}_x[m]$[3], such that the regularization term can be computed as

$$\underline{\mathbf{r}}_{2NL}[m] = \frac{1}{\mu}\mathbf{Q}_x[m]\,\underline{\mathbf{w}}_{2NL}[m-1]. \tag{1.67}$$

In fact, using the correlation matrix $\mathbf{Q}_x[m]$ instead of $\mathbf{D}_x^H[m]\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_x[m]$ is quite similar to the low-pass filtering of the time-domain regularization term, which has been proposed in [26] to improve the performance in highly non-stationary noise. In practice, using the assumption that speech and noise are uncorrelated, the speech correlation matrix is approximated as

$$\mathbf{Q}_x[m] = \mathbf{Q}_y[m] - \mathbf{Q}_v[m], \tag{1.68}$$

where $\mathbf{Q}_y[m]$ is the $2NL\times 2NL$-dimensional correlation matrix updated during speech-periods, i.e.

$$\mathbf{Q}_y[m] = \lambda\mathbf{Q}_y[m-1] + (1-\lambda)\,\mathbf{D}_y^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_y[m], \tag{1.69}$$

with $\mathbf{D}_y[m]$ defined similarly as $\mathbf{D}_x[m]$. For a speech source at a fixed position, it may also be possible to determine $\mathbf{Q}_x[m]$ using calibration signals [43,44] or by additionally incorporating geometric constraints [45,46]. In conclusion, the total frequency-domain adaptive algorithm for the SDW-MWF is summarized in Algorithm 1.

[3] Note that a similar reasoning for computing the term $\mathbf{D}_v^H[m]\,\underline{\mathbf{e}}_{v,2L}[m]$ during speech-periods is not possible, since

$$\mathbf{D}_v^H[m]\,\underline{\mathbf{e}}_{v,2L}[m] = \mathbf{D}_v^H[m]\,\underline{\mathbf{d}}_{2L}[m] - \mathbf{D}_v^H[m]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_v[m]\,\underline{\mathbf{w}}_{2NL}[m-1] \tag{1.66}$$

cannot be easily approximated, because of the term $\mathbf{D}_v^H[m]\,\underline{\mathbf{d}}_{2L}[m]$.

Matrix definitions:

$\mathbf{F}_{2L} = 2L \times 2L$-dimensional DFT matrix

$\mathbf{0}_L = L \times L$-dimensional zero matrix, $\quad \mathbf{I}_L = L \times L$-dimensional identity matrix

$$\mathbf{G}^{01}_{2L\times 2L} = \mathbf{F}_{2L} \begin{bmatrix} \mathbf{0}_L & \mathbf{0}_L \\ \mathbf{0}_L & \mathbf{I}_L \end{bmatrix} \mathbf{F}^{-1}_{2L}, \quad \mathbf{G}^{10}_{2L\times 2L} = \mathbf{F}_{2L} \begin{bmatrix} \mathbf{I}_L & \mathbf{0}_L \\ \mathbf{0}_L & \mathbf{0}_L \end{bmatrix} \mathbf{F}^{-1}_{2L}$$

$$\mathbf{G}^{10}_{2NL\times 2NL} = \mathrm{diag}\left[\, \mathbf{G}^{10}_{2L\times 2L} \ \ldots \ \mathbf{G}^{10}_{2L\times 2L} \,\right]$$

For each new block of L samples:

$$\mathbf{d}[m] = \left[\, y_0[mL - \Delta] \ y_0[mL - \Delta + 1] \ \ldots \ y_0[mL - \Delta + L - 1] \,\right]^T$$

$$\mathbf{D}_{y,n}[m] = \mathrm{diag}\left\{ \mathbf{F}_{2L} \left[\, y_n[mL - L] \ \ldots \ y_n[mL + L - 1] \,\right]^T \right\}, \qquad n = M - N \ldots M - 1$$

$$\mathbf{D}_y[m] = \left[\, \mathbf{D}_{y,M-N}[m] \ \ldots \ \mathbf{D}_{y,M-1}[m] \,\right]$$

Output signal:

$$\mathbf{e}[m] = \mathbf{d}[m] - \left[\, \mathbf{0}_L \ \mathbf{I}_L \,\right] \mathbf{F}^{-1}_{2L}\, \mathbf{D}_y[m]\, \underline{\mathbf{w}}_{2NL}[m-1]$$

If speech detected:

$$\mathbf{Q}_y[m] = \lambda \mathbf{Q}_y[m-1] + (1-\lambda)\, \mathbf{D}^H_y[m]\, \mathbf{G}^{01}_{2L\times 2L} \mathbf{D}_y[m]$$

If noise detected: $\mathbf{D}_v[m] = \mathbf{D}_y[m]$

$$\mathbf{Q}_v[m] = \lambda \mathbf{Q}_v[m-1] + (1-\lambda)\, \mathbf{D}^H_v[m]\, \mathbf{G}^{01}_{2L\times 2L} \mathbf{D}_v[m]$$

Update formula (only during noise-only-periods):

$$\underline{\mathbf{e}}_{v,2L}[m] = \mathbf{F}_{2L} \begin{bmatrix} \mathbf{0}_L \\ \mathbf{I}_L \end{bmatrix} \mathbf{e}[m]$$

$$\mathbf{Q}_x[m] = \mathbf{Q}_y[m] - \mathbf{Q}_v[m]$$

$$\underline{\mathbf{r}}_{2NL}[m] = \tfrac{1}{\mu} \mathbf{Q}_x[m]\, \underline{\mathbf{w}}_{2NL}[m-1]$$

$$\underline{\mathbf{w}}_{2NL}[m] = \underline{\mathbf{w}}_{2NL}[m-1] + (1-\lambda)\, \mathbf{G}^{10}_{2NL\times 2NL} \left[\mathbf{Q}_v[m] + \tfrac{1}{\mu} \mathbf{Q}_x[m] \right]^{-1} \cdot \left\{ \mathbf{D}^H_v[m]\, \underline{\mathbf{e}}_{v,2L}[m] - \underline{\mathbf{r}}_{2NL}[m] \right\}$$

Algorithm 1: Frequency-domain implementation of SDW-MWF.

1.4 Approximations for Reducing the Complexity

Algorithm 1 presents a general framework from which different adaptive algorithms can be derived. Some of these algorithms have already been presented in the literature [27,33], while other algorithms represent novel techniques for implementing the speech distortion weighted multichannel Wiener filter in the frequency domain.

1.4.1 Block-Diagonal Correlation Matrices

Since the correlation matrices $\mathbf{Q}_v[m]$ and $\mathbf{Q}_y[m]$ have no special structure, the update of these correlation matrices according to (1.51) and (1.69), and the matrix inverse required in (1.63) are computationally quite complex operations, such that in fact Algorithm 1 is not very useful in practice. However, in [34–36] it has been shown that the matrix $\mathbf{G}^{01}_{2L\times 2L}$ may be well approximated by $\mathbf{I}_{2L}/2$, because – for large L – the off-diagonal elements of $\mathbf{G}^{01}_{2L\times 2L}$ are small compared to its diagonal elements.

When using this approximation, we obtain the following update formula for the block-diagonal correlation matrices $\tilde{\mathbf{Q}}_v[m]$ and $\tilde{\mathbf{Q}}_y[m]$,

$$\tilde{\mathbf{Q}}_v[m] = \lambda\tilde{\mathbf{Q}}_v[m] + (1-\lambda)\,\mathbf{D}_v^H[m]\,\mathbf{D}_v[m]/2, \tag{1.70}$$

$$\tilde{\mathbf{Q}}_y[m] = \lambda\tilde{\mathbf{Q}}_y[m] + (1-\lambda)\,\mathbf{D}_y^H[m]\,\mathbf{D}_y[m]/2, \tag{1.71}$$

which consist of N^2 $2L\times 2L$-dimensional diagonal sub-matrices $\tilde{\mathbf{Q}}_{v,np}[m]$ and $\tilde{\mathbf{Q}}_{y,np}[m]$, $n = M-N\ldots M-1$, $p = M-N\ldots M-1$. In addition, we obtain the following update formula for the filter coefficients,

$$\underline{\mathbf{w}}_{2NL}[m] = \underline{\mathbf{w}}_{2NL}[m-1] + \rho(1-\lambda)\,\mathbf{G}^{10}_{2NL\times 2NL}\left[\tilde{\mathbf{Q}}_v[m] + \frac{1}{\mu}\tilde{\mathbf{Q}}_x[m]\right]^{-1}\cdot \left\{\mathbf{D}_v^H[m]\,\underline{\mathbf{e}}_{v,2L}[m] - \underline{\mathbf{r}}_{2NL}[m]\right\}, \tag{1.72}$$

where the step size parameter ρ is assumed to be in the range $0 < \rho \leq 1$. This equation requires the computation of the inverse of the block-diagonal matrix $\tilde{\mathbf{Q}}_v[m] + 1/\mu\,\tilde{\mathbf{Q}}_x[m]$. It is well known that the inverse of a block-diagonal matrix $\mathbf{Q}$, consisting of N^2 $2L\times 2L$-dimensional diagonal sub-matrices $\mathbf{Q}_{np}$, i.e.

$$\mathbf{Q} = \begin{bmatrix} \mathbf{Q}_{M-N,M-N} & \cdots & \mathbf{Q}_{M-N,M-1} \\ \vdots & & \vdots \\ \mathbf{Q}_{M-1,M-N} & \cdots & \mathbf{Q}_{M-1,M-1} \end{bmatrix}, \tag{1.73}$$

is also a block-diagonal matrix. Its computation corresponds to inverting $2L$ $N\times N$-dimensional matrices, which is attractive from a complexity point of view. The block-diagonal matrix $\mathbf{Q}$ can be easily transformed to the diagonal block-matrix $\bar{\mathbf{Q}}$,

$$\bar{\mathbf{Q}} = \mathrm{diag}\left[\,\bar{\mathbf{Q}}_0\ \ldots\ \bar{\mathbf{Q}}_{2L-1}\,\right], \tag{1.74}$$

consisting of $2L$ $N\times N$-dimensional sub-matrices $\bar{\mathbf{Q}}_l$, $l = 0\ldots 2L-1$, on its diagonal, by the transformation

$$\bar{\mathbf{Q}} = \mathbf{A}^T\,\mathbf{Q}\,\mathbf{A}. \tag{1.75}$$

The matrix $\mathbf{A}$ is a $2NL \times 2NL$-dimensional column permutation matrix (and hence $\mathbf{A}^T$ is a row permutation matrix), consisting of $2NL$ $2L \times N$-dimensional sub-matrices $\mathbf{A}_{nl}$, $n = M - N \dots M - 1$, $l = 0 \dots 2L - 1$, where the (l, n)-th element of $\mathbf{A}_{nl}$ is equal to 1. It readily follows that

$$\mathbf{Q}^{-1} = \mathbf{A}\, \bar{\mathbf{Q}}^{-1}\, \mathbf{A}^T, \tag{1.76}$$

where $\bar{\mathbf{Q}}^{-1}$ can be easily computed by inverting the $N \times N$-dimensional sub-matrices $\bar{\mathbf{Q}}_l$ on its diagonal, i.e.

$$\bar{\mathbf{Q}}^{-1} = \mathrm{diag}\left[\, \bar{\mathbf{Q}}_0^{-1} \dots \bar{\mathbf{Q}}_{2L-1}^{-1} \,\right]. \tag{1.77}$$

In addition, one should take care that the matrix $\tilde{\mathbf{Q}}_v[m] + 1/\mu\, \tilde{\mathbf{Q}}_x[m]$ in (1.72) is positive definite. When this matrix is not positive definite, this actually has the same effect as a negative step size ρ, i.e. divergence of the filter coefficients. The noise correlation matrix $\tilde{\mathbf{Q}}_v[m]$ is always positive definite, but the speech correlation matrix $\tilde{\mathbf{Q}}_x[m]$ may not always be positive definite (especially for non-stationary signals), since it is computed as $\tilde{\mathbf{Q}}_x[m] = \tilde{\mathbf{Q}}_y[m] - \tilde{\mathbf{Q}}_v[m]$, where $\tilde{\mathbf{Q}}_y[m]$ and $\tilde{\mathbf{Q}}_v[m]$ are estimated during (different) speech-periods and noise-only-periods. Checking the positive definiteness of a matrix comes down to computing its eigenvalues.

It can be easily shown that the eigenvalues γ of a block-diagonal matrix $\mathbf{Q}$ are equal to the set of eigenvalues of its $N \times N$-dimensional sub-matrices $\bar{\mathbf{Q}}_l$, $l = 0 \dots 2L - 1$, since using (1.75) and the fact that $\mathbf{A}^T\mathbf{A} = \mathbf{I}_{2NL}$ and $\det(\mathbf{A}) = \pm 1$, it follows that

$$\det(\mathbf{Q} - \gamma \mathbf{I}_{2NL}) = \det\left(\mathbf{A}(\bar{\mathbf{Q}} - \gamma \mathbf{I}_{2NL})\mathbf{A}^T\right) = \det(\bar{\mathbf{Q}} - \gamma \mathbf{I}_{2NL}) \tag{1.78}$$

$$= \prod_{l=0}^{2L-1} \det(\bar{\mathbf{Q}}_l - \gamma \mathbf{I}_N). \tag{1.79}$$

Hence, before computing the inverse of the matrix $\tilde{\mathbf{Q}}_v[m] + 1/\mu\, \tilde{\mathbf{Q}}_x[m]$ in (1.72), we first compute the eigenvalues of the matrix $\tilde{\mathbf{Q}}_x[m]$. We will then use the inverse of the regularized matrix

$$\tilde{\mathbf{Q}}_v[m] + \frac{1}{\mu}\, \left[\tilde{\mathbf{Q}}_x[m] - \min(\gamma_{min}, 0)\, \mathbf{I}_{2NL}\right] + \delta\, \mathbf{I}_{2NL} \tag{1.80}$$

in (1.72), with γ_{min} the smallest eigenvalue of $\tilde{\mathbf{Q}}_x[m]$ and δ a small positive regularization factor. Whereas in general computing the smallest eigenvalue of an $N \times N$-dimensional Hermitian matrix is computationally quite complex, for $N = 2$ the smallest eigenvalue $\gamma_{l,min}$ of the sub-matrix

$$\bar{\mathbf{Q}}_l = \begin{bmatrix} \bar{q}_{l,11} & \bar{q}_{l,12} \\ \bar{q}_{l,12}^* & \bar{q}_{l,22} \end{bmatrix}, \tag{1.81}$$

with $\bar{q}_{l,11}$ and $\bar{q}_{l,22}$ real-valued, can be easily computed as

$$\gamma_{l,min} = \frac{(\bar{q}_{l,11} + \bar{q}_{l,22}) - \sqrt{(\bar{q}_{l,11} - \bar{q}_{l,22})^2 + 4|\bar{q}_{l,12}|^2}}{2}. \tag{1.82}$$

1.4.2 Diagonal Correlation Matrices

In a further approximation, we can decouple the updates for the N filters $\underline{\mathbf{w}}_{n,2L}[m]$ in (1.72) by neglecting the off-diagonal elements of the matrix $\tilde{\mathbf{Q}}_v[m] + 1/\mu\,\tilde{\mathbf{Q}}_x[m]$, which represent the inter-channel correlation. Hence, the update formula for the filter coefficients $\underline{\mathbf{w}}_{n,2L}[m]$, $n = M-N \ldots M-1$ becomes

$$\underline{\mathbf{w}}_{n,2L}[m] = \underline{\mathbf{w}}_{n,2L}[m-1] + \rho(1-\lambda)\,\mathbf{G}^{10}_{2L\times 2L}\left[\tilde{\mathbf{Q}}_{v,nn}[m] + \frac{1}{\mu}\tilde{\mathbf{Q}}_{x,nn}[m]\right]^{-1} \cdot \left\{\mathbf{D}^H_{v,n}[m]\,\underline{\mathbf{e}}_{v,2L}[m] - \underline{\mathbf{r}}_{n,2L}[m]\right\}, \tag{1.83}$$

with $\tilde{\mathbf{Q}}_{v,nn}[m]$ and $\tilde{\mathbf{Q}}_{x,nn}[m]$ the $2L \times 2L$-dimensional diagonal sub-matrices on the diagonal of $\tilde{\mathbf{Q}}_v[m]$ and $\tilde{\mathbf{Q}}_x[m]$, and $\underline{\mathbf{r}}_{n,2L}[m]$ a $2L$-dimensional sub-vector of $\underline{\mathbf{r}}_{2NL}[m]$[4]. Ensuring the positive definiteness of $\tilde{\mathbf{Q}}_{x,nn}[m]$ now is very easy, since the eigenvalues of $\tilde{\mathbf{Q}}_{x,nn}[m]$ are equal to the diagonal elements. We expect that updating the filter coefficients using these diagonal correlation matrices will be slower than using the block-diagonal correlation matrices, since the inter-channel correlation is not taken into account any more.

Where in (1.83) a different step size matrix $\tilde{\mathbf{Q}}_{v,nn}[m] + 1/\mu\tilde{\mathbf{Q}}_{x,nn}[m]$ is used for each channel n, it is also possible to use a common step size matrix $\tilde{\mathbf{Q}}_c$, e.g. the sum or average over all channels, i.e.

$$\underline{\mathbf{w}}_{n,2L}[m] = \underline{\mathbf{w}}_{n,2L}[m-1] + \rho(1-\lambda)\,\mathbf{G}^{10}_{2L\times 2L}\tilde{\mathbf{Q}}_c^{-1}[m] \cdot \left\{\mathbf{D}^H_{v,n}[m]\,\underline{\mathbf{e}}_{v,2L}[m] - \underline{\mathbf{r}}_{n,2L}[m]\right\}, \tag{1.84}$$

$$\tilde{\mathbf{Q}}_c[m] = \left(\frac{1}{N}\right)\sum_{n=M-N}^{M-1}\tilde{\mathbf{Q}}_{v,nn}[m] + \frac{1}{\mu}\tilde{\mathbf{Q}}_{x,nn}[m]. \tag{1.85}$$

In fact, this algorithm is very similar to the algorithm already presented in [27]. Note however that the algorithm in [27] has been derived as a frequency-domain implementation of a time-domain stochastic gradient algorithm for minimizing the (time-domain) cost function (1.14).

1.4.3 Unconstrained Algorithms

In Sect. 1.4.1 the term $\mathbf{G}^{01}_{2L\times 2L}$ in the calculation of the correlation matrices has been approximated by $\mathbf{I}_{2L}/2$. It is also possible to use the same approximation for the term $\mathbf{G}^{10}_{2L\times 2L}$ and hence approximate $\mathbf{G}^{10}_{2NL\times 2NL}$ in the update formula of the filter coefficients in (1.63) by

$$\mathbf{G}^{10}_{2NL\times 2NL} \approx \mathrm{diag}\left[\,\mathbf{I}_{2L}/2 \ldots \mathbf{I}_{2L}/2\,\right] = \mathbf{I}_{2NL}/2, \tag{1.86}$$

[4] Note that we still use the off-diagonal elements of $\tilde{\mathbf{Q}}_x[m]$ for computing the regularization term $\underline{\mathbf{r}}_{2NL}[m]$.

resulting in the following so-called unconstrained update formula,

$$\underline{\mathbf{w}}_{2NL}[m] = \underline{\mathbf{w}}_{2NL}[m-1] + \frac{(1-\lambda)}{2}\left[\mathbf{Q}_v[m] + \frac{1}{\mu}\mathbf{Q}_x[m]\right]^{-1} \cdot \left\{\mathbf{D}_v^H[m]\,\underline{\mathbf{e}}_{v,2L}[m] - \underline{\mathbf{r}}_{2NL}[m]\right\}. \tag{1.87}$$

This update formula gives rise to a lower computational complexity, since it requires $2N$ less FFT operations. However, using this update formula it cannot be guaranteed that the second half of $\mathbf{F}_{2L}^{-1}\underline{\mathbf{w}}_{n,2L}[m]$, $n = M-N \ldots M-1$, is equal to zero, cf. (1.55).

In fact, this update formula can also be derived by setting the derivative of the cost function $J_f[m]$ in (1.35) with respect to the (frequency-domain) filter coefficients $\underline{\mathbf{w}}_{2NL}[m]$ equal to zero. Since from (1.32), (1.33), (1.46) and (1.47), it readily follows that $\underline{\mathbf{e}}_{v,2L}^H[i]\,\underline{\mathbf{e}}_{v,2L}[i] = 2\,\underline{\mathbf{e}}_v^H[i]\,\underline{\mathbf{e}}_v[i]$ and $\underline{\mathbf{e}}_{x,2L}^H[i]\,\underline{\mathbf{e}}_{x,2L}[i] = 2\,\underline{\mathbf{e}}_x^H[i]\,\underline{\mathbf{e}}_x[i]$, the cost function $J_f[m]$ can be written as

$$J_f[m] = \frac{(1-\lambda_v)}{2}\sum_{i=0}^{m}\lambda_v^{m-i}\underline{\mathbf{e}}_{v,2L}^H[i]\underline{\mathbf{e}}_{v,2L}[i] + \frac{1}{\mu}\frac{(1-\lambda_x)}{2}\sum_{i=0}^{m}\lambda_x^{m-i}\underline{\mathbf{e}}_{x,2L}^H[i]\underline{\mathbf{e}}_{x,2L}[i].$$

Hence, using (1.56) and (1.57) and the fact that $(\mathbf{G}_{2L\times 2L}^{01})^H\mathbf{G}_{2L\times 2L}^{01} = \mathbf{G}_{2L\times 2L}^{01}$ and $(\mathbf{G}_{2L\times 2L}^{01})^H\underline{\mathbf{d}}_{2L}[i] = \underline{\mathbf{d}}_{2L}[i]$, the derivative of $J_f[m]$ with respect to $\underline{\mathbf{w}}_{2NL}[m]$ is equal to

$$\frac{\partial J_f[m]}{\partial \underline{\mathbf{w}}_{2NL}[m]} = (1-\lambda_v)\sum_{i=0}^{m}\lambda_v^{m-i}\Big(\mathbf{D}_v^H[i]\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_v[i]\underline{\mathbf{w}}_{2NL}[m] - \mathbf{D}_v^H[i]\underline{\mathbf{d}}_{2L}[i]\Big) + \frac{1}{\mu}(1-\lambda_x)\sum_{i=0}^{m}\lambda_x^{m-i}\,\mathbf{D}_x^H[i]\,\mathbf{G}_{2L\times 2L}^{01}\mathbf{D}_x[i]\underline{\mathbf{w}}_{2NL}[m]. \tag{1.88}$$

Setting the derivative equal to zero, one obtains the normal equations

$$\left[\mathbf{Q}_v[m] + \frac{1}{\mu}\mathbf{Q}_x[m]\right]\underline{\mathbf{w}}_{2NL}[m] = \mathbf{q}[m]\,, \tag{1.89}$$

with $\mathbf{Q}_v[m]$ and $\mathbf{Q}_x[m]$ defined in (1.51) and (1.52), and the $2NL$-dimensional vector $\mathbf{q}[m]$ equal to

$$\mathbf{q}[m] = \lambda_v\mathbf{q}[m-1] + (1-\lambda_v)\,\mathbf{D}_v^H[m]\,\underline{\mathbf{d}}_{2L}[m]. \tag{1.90}$$

Hence, by enforcing the normal equations at block time m and $m-1$ (and assuming $\lambda_x = \lambda_v = \lambda$), we obtain the update formula (1.87). In addition, for the unconstrained algorithms one can also approximate the correlation matrices $\mathbf{Q}_v[m]$ and $\mathbf{Q}_x[m]$ by block-diagonal or diagonal matrices, cf. Sects. 1.4.1 and 1.4.2.

Algorithm	Step size matrix
Algo 1 (C-BD)	$\mathbf{G}_{2NL\times 2NL}^{10}\left[\tilde{\mathbf{Q}}_v[m]+\frac{1}{\mu}\tilde{\mathbf{Q}}_x[m]\right]^{-1}$
Algo 2 (U-BD)	$\frac{1}{2}\left[\tilde{\mathbf{Q}}_v[m]+\frac{1}{\mu}\tilde{\mathbf{Q}}_x[m]\right]^{-1}$
Algo 3 (C-D1)	$\mathbf{G}_{2NL\times 2NL}^{10}\text{diag}\Big\{\left[\tilde{\mathbf{Q}}_{v,nn}[m]+\frac{1}{\mu}\tilde{\mathbf{Q}}_{x,nn}[m]\right]^{-1}\Big\}$
Algo 4 (U-D1)	$\frac{1}{2}\,\text{diag}\Big\{\left[\tilde{\mathbf{Q}}_{v,nn}[m]+\frac{1}{\mu}\tilde{\mathbf{Q}}_{x,nn}[m]\right]^{-1}\Big\}$
Algo 5 (C-D2)	$\mathbf{G}_{2NL\times 2NL}^{10}\text{diag}\Big\{\left[(1/N)\sum_{n=M-N}^{M-1}\tilde{\mathbf{Q}}_{v,nn}[m]+\frac{1}{\mu}\tilde{\mathbf{Q}}_{x,nn}[m]\right]^{-1}\Big\}$
Algo 6 (U-D2)	$\frac{1}{2}\,\text{diag}\Big\{\left[(1/N)\sum_{n=M-N}^{M-1}\tilde{\mathbf{Q}}_{v,nn}[m]+\frac{1}{\mu}\tilde{\mathbf{Q}}_{x,nn}[m]\right]^{-1}\Big\}$

Table 1.1. Step size matrix $\mathbf{\Lambda}[m]$ for different algorithms (C: constrained, U: unconstrained, BD: block-diagonal, D1: diagonal - channel, D2: diagonal - common).

1.4.4 Summary

Summarizing all presented algorithms in Sect. 1.4, the update formula for the filter coefficients $\underline{\mathbf{w}}_{2NL}[m]$ can be written as

$$\underline{\mathbf{w}}_{2NL}[m]=\underline{\mathbf{w}}_{2NL}[m-1]+\rho(1-\lambda)\,\mathbf{\Lambda}[m]\Big\{\mathbf{D}_v^H[m]\,\underline{\mathbf{e}}_{v,2L}[m]-\underline{\mathbf{r}}_{2NL}[m]\Big\},\quad(1.91)$$

where the $2NL\times 2NL$-dimensional step size matrix $\mathbf{\Lambda}[m]$ is summarized in Table 1.1. For all algorithms, the matrix $\tilde{\mathbf{Q}}_x[m]$ needs to be regularized in order to ensure that it is positive definite. The algorithm presented in [27] corresponds to **Algo 5**, while in [33] a multichannel frequency-domain algorithm for speech enhancement has been presented that bears quite some similarities to **Algo 2**. However, in [33] the speech component instead of the noise component is estimated, and the parameter $1/\mu$ is only effectively applied to the step size matrix (i.e. no regularization term $\underline{\mathbf{r}}_{2NL}[m]$ is present), such that the algorithm in [33] comes down to the MMSE estimation of the speech component.

1.5 Experimental Results

In this section experimental results are presented for a hearing aid application. For small-sized microphone arrays as typically used in hearing aids, robustness is very important, since small-sized microphone arrays exhibit a large sensitivity to signal model errors [47]. Sect. 1.5.1 describes the setup and defines the used performance measures. In Sect. 1.5.2 the performance of the adaptive algorithms is analyzed, and the impact of different parameter settings for the SP-SDW-MWF (i.e. filter $\mathbf{w}_0$ and $1/\mu$) on the performance and the robustness against signal model errors is evaluated. In Sect. 1.5.3 the tracking performance is investigated for a changing noise scenario.

1.5.1 Setup and Performance Measures

A behind-the-ear hearing aid with $M = 3$ omni-directional microphones (Knowles FG-3452) in an end-fire configuration has been mounted on a dummy head in an office room. The inter-spacing between the first and the second microphone is about 1 cm and the inter-spacing between the second and the third microphone is about 1.5 cm. The reverberation time $T_{60\text{dB}}$ of the office room is about 700 ms. The speech source and the noise sources are positioned at a distance of 1 m from the head: the speech source in front of the head (0°), and the noise sources at an angle θ with respect to the speech source. Both the speech signal and the noise signal have a level of 70 dB at the center of the head. For evaluation purposes, the speech and the noise signal have been recorded separately. The sampling frequency is equal to 16 kHz.

The microphone signals are pre-whitened prior to processing in order to improve the intelligibility, and the output signal $z[k]$ is accordingly de-whitened [48]. In the experiments, the microphones have been calibrated using anechoic recordings of a speech-weighted noise signal at 0° with the microphone array mounted on the head. A delay-and-sum beamformer is used for the fixed beamformer $\mathbf{A}(z)$, since – in the case of small microphone inter-spacing – this beamformer is quite robust to signal model errors. The blocking matrix $\mathbf{B}(z)$ pair-wisely subtracts the time-aligned calibrated microphone signals.

To assess the performance of the different algorithms, the broadband intelligibility weighted signal-to-noise ratio improvement $\Delta\text{SNR}_{\text{intellig}}$ is used, which is defined as [49]

$$\Delta\text{SNR}_{\text{intellig}} = \sum_i I_i \left(\text{SNR}_{i,\text{out}} - \text{SNR}_{i,\text{in}}\right), \tag{1.92}$$

where the band importance function I_i expresses the importance of the ith one-third octave band with center frequency f_i^c for intelligibility, and where $\text{SNR}_{i,\text{out}}$ and $\text{SNR}_{i,\text{in}}$ are respectively the output SNR and the input SNR (in dB) in this band. The center frequencies f_i^c and the values I_i are defined in [50]. The intelligibility weighted SNR improvement reflects how much the speech intelligibility is improved by the noise reduction algorithms, but does not take into account speech distortion.

In order to measure the amount of (linear) speech distortion, we similarly define an intelligibility weighted spectral distortion measure $\text{SD}_{\text{intellig}}$,

$$\text{SD}_{\text{intellig}} = \sum_i I_i \, \text{SD}_i \ , \tag{1.93}$$

with SD_i the average spectral distortion (dB) in the ith one-third octave band, calculated as

$$\text{SD}_i = \frac{1}{\left(2^{1/6} - 2^{-1/6}\right) f_i^c} \int_{2^{-1/6} f_i^c}^{2^{1/6} f_i^c} \left|10 \log_{10} G_x(f)\right| df, \tag{1.94}$$

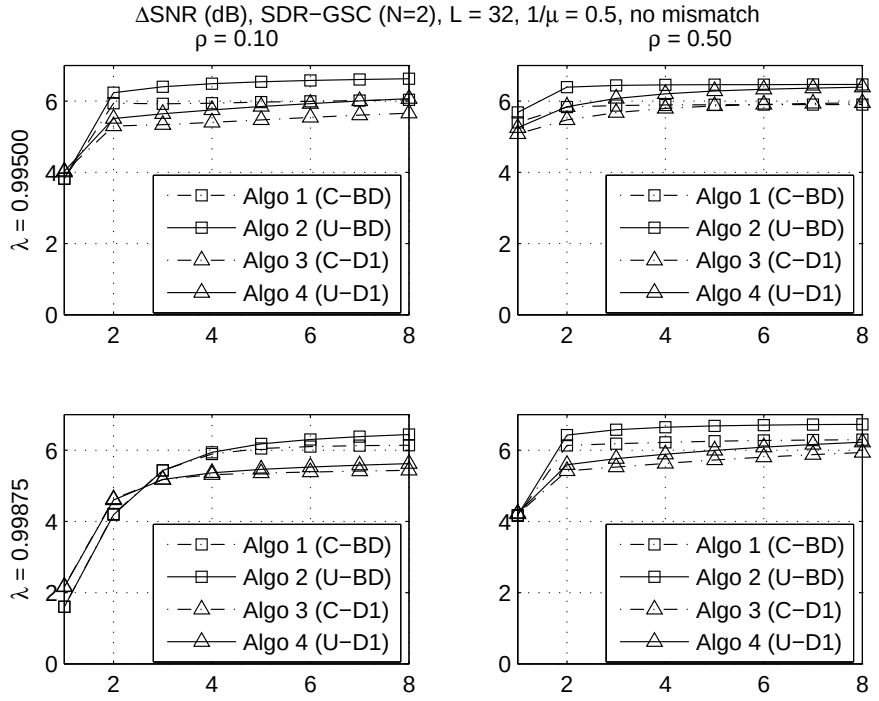

Fig. 1.2. Effect of the step size parameter ρ and the exponential forgetting factor λ on the convergence of the SNR improvement for different adaptive algorithms (SDR-GSC, $L = 32$, $1/\mu = 0.5$, no microphone mismatch).

with $G_x(f)$ the power transfer function for the speech component from the input to the output of the noise reduction algorithm.

All algorithms are evaluated for a filter length $L = 32$, and the input SNR of the microphone signals is equal to 0 dB. The speech-periods and the noise-only-periods, used for updating the correlation matrices $\tilde{\mathbf{Q}}_y[m]$ and $\tilde{\mathbf{Q}}_v[m]$ and the adaptive filter, have been marked manually. In order to exclude the effect of the spatial pre-processor, the performance measures (1.92) and (1.93) are calculated with respect to the output of the fixed beamformer, i.e. the speech reference $y_0[k]$. In some experiments, a microphone gain mismatch of 4 dB is applied to the second microphone to illustrative the sensitivity to signal model errors. Among the different possible signal model errors, microphone mismatch has been found to be quite harmful to the performance of the GSC in a hearing aid application [47]. In hearing aids, microphones are rarely matched in gain and phase, with gain and phase differences between microphone characteristics of up to 6 dB and 10° [51].

1.5.2 SNR Improvement and Robustness Against Microphone Mismatch

For the experiments in this section, the desired speech source at 0° consists of sentences from the HINT-database [52] spoken by a male speaker, and a complex noise scenario consisting of 5 spectrally non-stationary multi-talker babble noise sources at 75°, 120°, 180°, 240° and 285°, has been used.

Figure 1.2 depicts the convergence of the SNR improvement for different adaptive algorithms (constrained vs. unconstrained, block-diagonal vs. diagonal step size matrix) for different values of the the step size parameter ρ and the exponential forgetting factor λ. The exponential forgetting factor $\lambda = 0.995$ corresponds to an averaging of the correlation matrices over approximately $1/(1-\lambda) = 200$ blocks of $L = 32$ samples, i.e. 0.4 seconds, whereas the factor $\lambda = 0.99875$ corresponds to an averaging over 800 blocks, i.e. 1.6 seconds. In this experiment, we have used the SDR-GSC ($N = 2$) with trade-off parameter $1/\mu = 0.5$ and no microphone mismatch present. Of course, similar plots can be obtained for the SP-SDW-MWF ($N = 3$), for different values of the trade-off parameter and when microphone mismatch is present. From Fig. 1.2 it can be seen that using a block-diagonal step size matrix gives rise to a faster convergence than using a diagonal step size matrix. In addition, the larger the step size parameter ρ, the faster the convergence (of course, taking ρ too large will give rise to divergence). ¿From Fig. 1.2 it can also be seen that the larger the exponential forgetting factor λ, the slower the convergence, but the larger the SNR improvement after convergence. This can be explained by the fact that for spectrally and/or spatially stationary sources a better estimate of the correlation matrices is obtained for larger λ.

Figure 1.3 plots the SNR improvement and the speech distortion of the SDR-GSC ($N = 2$), using the unconstrained update formula (with block-diagonal and diagonal step size matrix), as a function of the trade-off parameter $1/\mu$. This figure also depicts the effect of a gain mismatch of 4 dB at the second microphone. In the absence of microphone mismatch, the amount of speech leakage into the noise references is limited, such that the speech distortion is small for all $1/\mu$. However, since there is some speech leakage present due to reverberation, the SNR improvement decreases for increasing $1/\mu$. In the presence of microphone mismatch, the amount of speech leakage into the noise references grows. For the standard GSC, i.e. $1/\mu = 0$, significant speech distortion now occurs and the SNR improvement is seriously degraded. Setting $1/\mu > 0$ improves the performance of the GSC in the presence of signal model errors, i.e. speech distortion decreases and the SNR degradation becomes smaller. For the given setup, a value $1/\mu = 0.5$ seems appropriate for guaranteeing good performance for a gain mismatch up to 4 dB.

For the same setup, Fig. 1.4 plots the SNR improvement and the speech distortion of the SP-SDW-MWF ($N = 3$) as a function of $1/\mu$. This figure shows that the speech distortion and the SNR improvement decreases

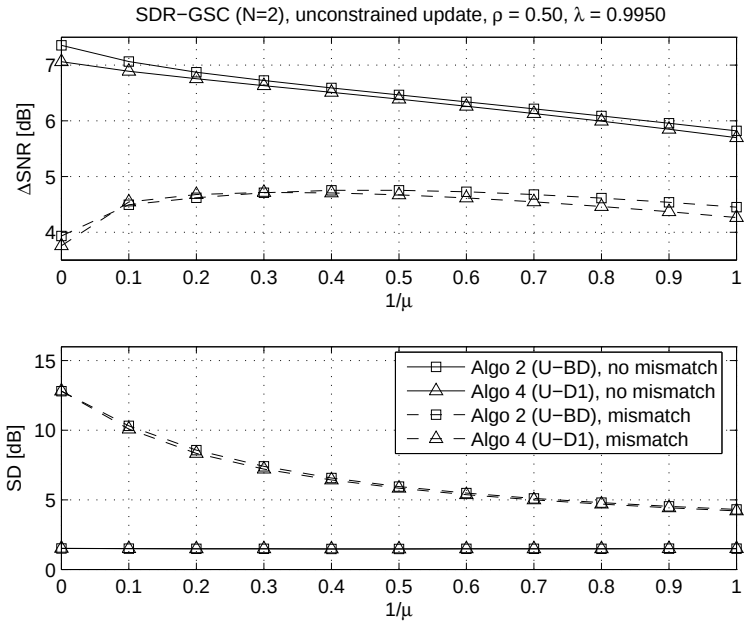

Fig. 1.3. SNR improvement and speech distortion of SDR-GSC in function of trade-off parameter $1/\mu$, without and with gain mismatch (unconstrained update formula, $\rho = 0.5$, $\lambda = 0.995$).

for increasing $1/\mu$. This figure also shows that the speech distortion for the SP-SDW-MWF is larger than for the SDR-GSC[5], but that both the SNR improvement and the speech distortion are hardly affected by microphone mismatch.

1.5.3 Tracking Performance

In order to investigate the tracking performance, we have used a stationary speech-weighted noise signal both for the desired speech source and for the noise sources. We consider 2 noise scenarios: scenario 1 with five noise sources at 75°, 120°, 180°, 240° and 285°, and scenario 2 with three noise sources at 120°, 240° and 285°. The noise scenario suddenly changes from scenario 1 to scenario 2 after 45 seconds, and then changes back to scenario 1 after 90 seconds. The speech component consists of alternating segments of silence and signal, each with a length of 1600 samples. For both noise scenarios, the signals received at the microphones have been normalized, such that for both scenarios the input SNR is equal to 0 dB.

[5] In [25], it has been shown that the SP-SDW-MWF can be interpreted as an SDR-GSC with a single-channel post-filter in the absence of speech leakage.

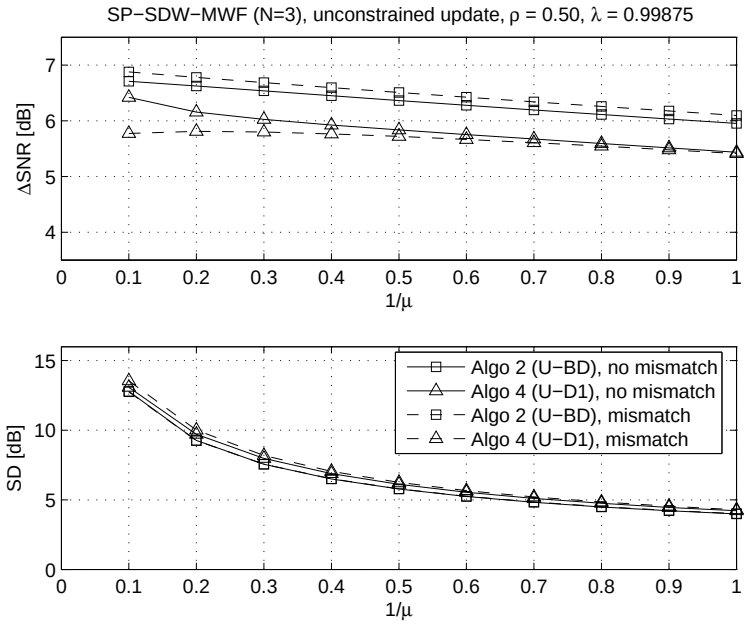

Fig. 1.4. SNR improvement and speech distortion of SP-SDW-MWF in function of trade-off parameter $1/\mu$, without and with gain mismatch (unconstrained update formula, $\rho = 0.5$, $\lambda = 0.99875$).

Figure 1.5 plots the SNR improvement, the speech distortion and the residual noise energy ε_v^2 for the GSC ($1/\mu = 0$) and the SDR-GSC ($1/\mu = 0.5$) using the unconstrained update formula with block-diagonal step size matrix (Algo 2), in the case of a gain mismatch of 4 dB. These performance measures have been calculated per segment of 3200 samples. Again, this figure shows that the SDR-GSC is more robust to signal model errors than the GSC, since the SDR-GSC gives rise to a larger SNR improvement and a smaller speech distortion than the GSC, although the residual noise energy is larger (for both noise scenarios). This figure also shows that the performance of the SDR-GSC for noise scenario 2 is better than for noise scenario 1, and that the SDR-GSC is able to track sudden changes in noise scenarios.

1.6 Conclusions

In this chapter, we have discussed a robust multi-microphone speech enhancement technique, called the spatially pre-processed speech distortion weighted multichannel Wiener filter (SP-SDW-MWF). The SP-SDW-MWF takes speech distortion due to speech leakage explicitly into account in the design criterion of the adaptive filter, and is hence more robust against signal

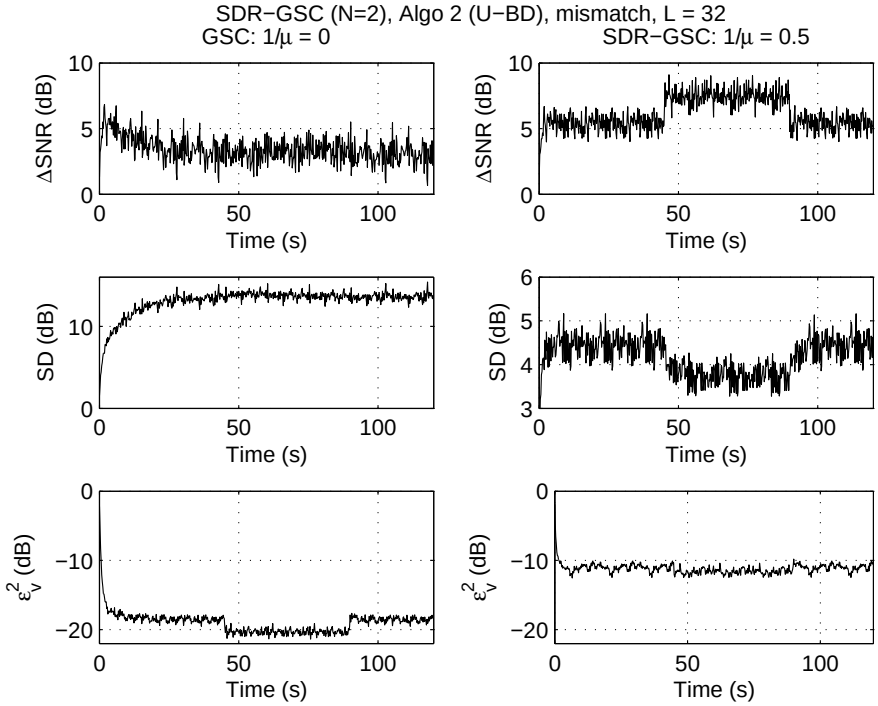

Fig. 1.5. Tracking performance of GSC and SDR-GSC for a changing noise scenario.

model errors than the standard GSC. Depending on its parameter setting, the SP-SDW-MWF encompasses the standard GSC and the speech distortion regularized GSC (SDR-GSC) as special cases. We have presented a frequency-domain criterion for the SDW-MWF, which provides a way for linking existing adaptive frequency-domain algorithms and for deriving novel adaptive algorithms for implementing the SDW-MWF. The main difference between these adaptive algorithms consists in the calculation of the step size matrix (constrained vs. unconstrained, block-diagonal vs. diagonal) used in the update formula for the adaptive filter. Experimental results using a small-sized microphone array show that setting the trade-off parameter between noise reduction and speech distortion larger than 0 in the SDR-GSC improves the performance in the presence of signal model errors, i.e. the speech distortion decreases and the SNR degradation due to the signal model errors becomes smaller. Moreover, in the SP-SDW-MWF both the SNR improvement and the speech distortion are hardly affected by signal model errors. Experimental results also show that using a block-diagonal step size matrix gives rise to a faster convergence than using a diagonal step size matrix and that the SP-SDW-MWF is able to track changing noise scenarios.

References

1. B. D. Van Veen and K. M. Buckley, "Beamforming: a versatile approach to spatial filtering," *IEEE ASSP Magazine*, vol. 5, pp. 4–24, Apr. 1988.
2. O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. of the IEEE*, vol. 60, pp. 926–935, Aug. 1972.
3. L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. 30, pp. 27–34, Jan. 1982.
4. K. M. Buckley, "Broad-band beamforming and the generalized sidelobe canceller," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 34, pp. 1322–1323, Oct. 1986.
5. S. Nordholm, I. Claesson, and B. Bengtsson, "Adaptive array noise suppression of handsfree speaker input in cars," *IEEE Trans. Veh. Technol.*, vol. 42, pp. 514–518, Nov. 1993.
6. J. E. Greenberg and P. M. Zurek, "Evaluation of an adaptive beamforming method for hearing aids," *Journal of the Acoustical Society of America*, vol. 91, pp. 1662–1676, Mar. 1992.
7. J. Vanden Berghe and J. Wouters, "An adaptive noise canceller for hearing aids using two nearby microphones," *Journal of the Acoustical Society of America*, vol. 103, pp. 3621–3626, June 1998.
8. 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*, vol. 25, pp. 215–229, June 2004.
9. J. Bitzer, K. U. Simmer, and K.-D. Kammeyer, "Multi-microphone noise reduction techniques as front-end devices for speech recognition," *Speech Communication*, vol. 34, pp. 3–12, Apr. 2001.
10. W. Herbordt, H. Buchner, and W. Kellermann, "An acoustic human-machine front-end for multimedia applications," *EURASIP Journal on Applied Signal Processing*, vol. 2003, pp. 21–31, Jan. 2003.
11. I. Claesson and S. Nordholm, "A spatial filtering approach to robust adaptive beamforming," *IEEE Trans. Antennas Propagat.*, vol. 40, pp. 1093–1096, Sept. 1992.
12. S. Nordebo, I. Claesson, and S. Nordholm, "Adaptive beamforming: spatial filter designed blocking matrix," *IEEE Journal of Oceanic Engineering*, vol. 19, pp. 583–590, Oct. 1994.
13. S. Doclo and M. Moonen, "Design of broadband beamformers robust against gain and phase errors in the microphone array characteristics," *IEEE Trans. Signal Processing*, vol. 51, pp. 2511–2526, Oct. 2003.
14. D. Van Compernolle, "Switching adaptive filters for enhancing noisy and reverberant speech from microphone array recordings," in *Proc. IEEE ICASSP*, 1990, pp. 833–836.
15. 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*, vol. 47, pp. 2677–2684, Oct. 1999.
16. W. Herbordt and W. Kellermann, "Computationally efficient frequency-domain robust generalized sidelobe canceller," in *Proc. IWAENC*, 2001, pp. 51–54.

17. W. Herbordt and W. Kellermann, "Adaptive beamforming for audio signal acquisition," chapter 6 in *Adaptive Signal Processing: Applications to Real-World Problems*, J. Benesty and Y. Huang, Eds.), pp. 155–194, Springer-Verlag, 2003.
18. S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and non-stationarity with applications to speech," *IEEE Trans. Signal Processing*, vol. 49, pp. 1614–1626, Aug. 2001.
19. S. Gannot and I. Cohen, "Speech enhancement based on the general transfer function GSC and postfiltering," *IEEE Trans. Speech and Audio Processing*, vol. 12, pp. 561–571, Nov. 2004.
20. 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*, vol. E84-A, pp. 406–413, Feb. 2001.
21. N. K. Jablon, "Adaptive beamforming with the generalized sidelobe canceller in the presence of array imperfections," *IEEE Trans. Antennas Propagat.*, vol. 34, pp. 996–1012, Aug. 1986.
22. H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, pp. 1365–1376, Oct. 1987.
23. M. W. Hoffman and K. M. Buckley, "Robust time-domain processing of broadband microphone array data," *IEEE Trans. Speech and Audio Processing*, vol. 3, pp. 193–203, May 1995.
24. 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*, vol. 49, pp. 1138–1145, June 2001.
25. A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, pp. 2367–2387, Dec. 2004.
26. A. Spriet, M. Moonen, and J. Wouters, "Stochastic gradient implementation of spatially pre-processed multi-channel Wiener filtering for noise reduction in hearing aids," in *Proc. IEEE ICASSP*, 2004, pp. 57–60.
27. S. Doclo, A. Spriet, and M. Moonen, "Efficient frequency-domain implementation of speech distortion weighted multi-channel Wiener filtering for noise reduction," in *Proc. EUSIPCO*, 2004, pp. 2007–2010.
28. S. Doclo and M. Moonen, "GSVD-based optimal filtering for multi-microphone speech enhancement," chapter 6 in *Microphone Arrays: Signal Processing Techniques and Applications*, (M. S. Brandstein and D. B. Ward, Eds.), pp. 111–132, Springer-Verlag, 2001.
29. S. Doclo and M. Moonen, "GSVD-based optimal filtering for single and multi-microphone speech enhancement," *IEEE Trans. Signal Processing*, vol. 50, pp. 2230–2244, Sept. 2002.
30. A. Spriet, M. Moonen, and J. Wouters, "A multi-channel subband generalized singular value decomposition approach to speech enhancement," *European Transactions on Telecommunications, special issue on Acoustic Echo and Noise Control*, vol. 13, pp. 149–158, Mar.-Apr. 2002.
31. G. Rombouts and M. Moonen, "QRD-based unconstrained optimal filtering for acoustic noise reduction," *Signal Processing*, vol. 83, pp. 1889–1904, Sept. 2003.
32. D. A. Florêncio and H. S. Malvar, "Multichannel filtering for optimum noise reduction in microphone arrays," in *Proc. IEEE ICASSP*, 2001, pp. 197–200.

33. R. Aichner, W. Herbordt, H. Buchner, and W. Kellermann, "Least-squares error beamforming using minimum statistics and multichannel frequency-domain adaptive filtering," in *Proc. IWAENC*, 2003, pp. 223–226.
34. J. Benesty and D. R. Morgan, "Frequency-domain adaptive filtering revisited, generalization to the multi-channel case, and application to acoustic echo cancellation," in *Proc. IEEE ICASSP*, 2000, pp. 789–792.
35. J. Benesty, T. Gänsler, D. R. Morgan, M. M. Sondhi, and S. L. Gay, "General derivation of frequency-domain adaptive filtering," chapter 8 in *Advances in Network and Acoustic Echo Cancellation*, pp. 157–176, Springer-Verlag, 2001.
36. H. Buchner, J. Benesty, and W. Kellermann, "Multichannel frequency-domain adaptive filtering with application to multichannel acoustic echo cancellation," chapter 4 in *Adaptive Signal Processing: Applications to Real-World Problems* (J. Benesty and Y. Huang, Eds.), pp. 95–128, Springer-Verlag, 2003.
37. S. Van Gerven and F. Xie, "A comparative study of speech detection methods," in *Proc. EUROSPEECH*, 1997, vol. 3, pp. 1095–1098.
38. J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection," *IEEE Signal Processing Lett.*, vol. 6, pp. 1–3, Jan. 1999.
39. S. G. Tanyer and H. Özer, "Voice activity detection in nonstationary noise," *IEEE Trans. Speech and Audio Processing*, vol. 8, pp. 478–482, July 2000.
40. W. Herbordt, T. Trini, and W. Kellermann, "Robust spatial estimation of the signal-to-interference ratio for non-stationary mixtures," in *Proc. IWAENC*, 2003, pp. 247–250.
41. Y. Ephraim and H. L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, pp. 251–266, July 1995.
42. J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, pp. 14–37, Jan. 1992.
43. S. Nordholm, I. Claesson, and M. Dahl, "Adaptive microphone array employing calibration signals: an analytical evaluation," *IEEE Trans. Speech and Audio Processing*, vol. 7,pp. 241–252, May 1999.
44. S. Nordholm, I. Claesson, and N. Grbić, "Optimal and adaptive microphone arrays for speech input in automobiles," chapter 14 in *Microphone Arrays: Signal Processing Techniques and Applications*, (M. S. Brandstein and D. B. Ward, Eds.), pp. 307–330, Springer-Verlag, 2001.
45. H. Q. Dam, S. Nordholm, N. Grbić, and H. H. Dam, "Speech enhancement employing adaptive beamformer with recursively updated soft constraints," in *Proc. IWAENC*, 2003, pp. 307–310.
46. H. Q. Dam, S. Y. Low, H. H. Dam, and S. Nordholm, "Space constrained beamforming with source PSD updates," in *Proc. IEEE ICASSP*, 2004, pp. 93–96.
47. A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of GSVD based optimal filtering and generalized sidelobe canceller for hearing aid applications," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2001, pp. 31–34.
48. M. J. Link and K. M. Buckley, "Prewhitening for intelligibility gain in hearing aids arrays," *Journal of the Acoustical Society of America*, vol. 93, pp. 2139–2140, Apr. 1993.
49. 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*, vol. 94, pp. 3009–3010, Nov. 1993.

50. Acoustical Society of America, "ANSI S3.5-1997 American National Standard Methods for Calculation of the Speech Intelligibility Index," June 1997.
51. L. B. Jensen, "Hearing aid with adaptive matching of input transducers," United States Patent, no. 6,741,714, May 25, 2004.
52. 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*, vol. 95, pp. 1085–1099, Feb. 1994.