Universität Erlangen-Nürnberg

Chair of Multimedia Communications and Signal Processing

Prof. Dr.-Ing. Walter Kellermann

SIM Thesis

Noise Reduction by Multi-channel Wiener Filtering and Spatial Prediction

Lalitha Pakala

May 2011

Supervisor: Dipl.-Ing. Klaus Reindl

Declaration

I declare that the work is entirely my own, and was produced with no assistance from third parties. I certify that the work has not been submitted in the same or any similar form for assessment to any other examining body and all references direct and indirect are indicated as such and have been cited accordingly.

Erlangen, May 11th, 2011

Lalitha Pakala

Contents

List of Figures

Abstract

A more pronounced problem for hearing aid users is the inability to understand speech in noisy environments. This has led to the introduction of noise reduction algorithms in hearing aids. Among the single-channel and multi-channel noise reduction algorithms, multi-channel noise reduction algorithms seem to be interesting for hearing aid applications because of their high noise reduction performance.

In this thesis, two multi-channel noise reduction algorithms, which are known as, Multi-channel Wiener Filter (MWF) and Spatial Prediction (SP) are evaluated in the context of binaural hearing aids and their relation to the Minimum Variance Distortionless Response (MVDR) beamformer is discussed. It is well known that the MWF can be decomposed into an MVDR beamformer and a single-channel Wiener post filter. In this thesis, it will be shown that SP is an MVDR beamformer when only noise reduction is desired and no dereverberation is considered. With the MWF algorithm, some speech distortion is inevitably obtained whereas with the SP algorithm the constraint on the speech distortion error can be imposed to be zero. From the theoretical findings, the SP algorithm seems to be a better solution in the case of hearing aids, where the objective is to maintain the speech intelligibility. Theoretical findings are verified by experiments evaluated for different scenarios. From the experimental results, the SP algorithm seems to be appealing for hearing aids.

Chapter 1

Introduction

1.1 Motivation

Hearing impairment is an evergrowing problem throughout the world. It was estimated by Professor Adrian Davis of the British MRC Institute of Hearing Research that the total number of people suffering from hearing loss of more than 25 decibels (dB) will exceed 900 million by 2025 [1]. A hearing impaired person experiences many difficulties in the day to day communication, for example, when communicating via telephone or cellphone, in a teleconference system, or when watching television.

A hearing aid is a device that assists the hearing impaired to understand speech better and thereby improving their communicative ability. Hearing aids have been invented centuries ago. Starting from the horn-shaped trumpets till the modern digital hearing aids, they have been ever evolving with many improvements [14]. However, noise is overwhelming for hearing aid users and one of the main problems for the hearing aid users is the inability to perceive speech in noisy environments. The intelligibility of speech is reduced significantly in noisy environments. This has led to the introduction of the so-called noise reduction algorithms in the modern hearing aids, with the objective to keep up the quality and intelligibility of speech before being presented to the hearing aid user.

The objective of any noise reduction algorithm is to suppress the interfering signals and background noise as much as possible without distorting the desired signal, i.e., to improve the signal-to-noise ratio (SNR) while minimizing speech distortion, and consequently enhance speech intelligibility. The developed noise reduction algorithms can be categorized into single-channel and multi-channel algorithms based on the number of channels or microphones used.

Single-channel noise reduction algorithms can exploit only the spectral and temporal characteristics of the signal for noise and interference suppression. Therefore, their applications are limited in practical acoustic environments where reverberation and other ambient noise sources introduce also the spatial diversity. For example, in a multispeaker scenario such as the cocktail party, the speech and noise signals are considerably overlapped in time and frequency, which makes it difficult for single-channel algorithms to suppress noise without introducing speech distortion and artifacts. Therefore, single-channel algorithms do not seem to improve the speech intelligibility significantly [2].

When more than one microphone is available, the use of multiple microphones allow for the minimization of speech distortion and high noise reduction performance. The desired speech source and the noise sources are physically located at different positions. Multi-channel noise reduction algorithms exploit this spatial diversity of the speech and noise signals, in addition to the temporal and spectral characteristics. Therefore, multi-channel noise reduction algorithms are interesting for hearing aids as they can acheive a high SNR while minimizing the speech distortion simultaneously, thereby an improvement in the speech intelligibility can be acheived [2].

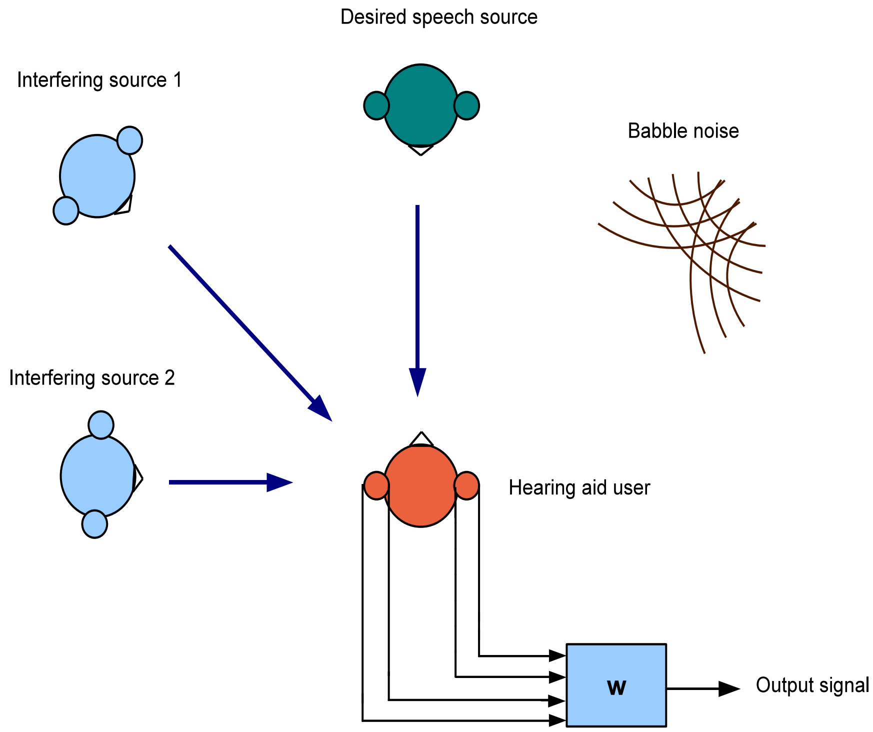


Figure 1.1: Signal acquisition scenario for a hearing aid user in a noisy environment

1.2 Initial problem

Let us consider the scenario as illustrated in Fig. 1.1, where a hearing aid user with hearing aids on both ears wants to listen to the desired speech source in the presence of noise and other interfering sources. Consider that each hearing aid is fitted with one or more microphones, in Fig. 1.1, the case with two microphones per hearing aid is shown. The received microphone signals do not only contain the desired speech signal but also noise and interfering signals. The objective is to estimate the desired speech signal contained in the noisy microphone signals.

In the following it is assumed that we have access to all the available microphone signals, i.e., the microphone signals from one hearing aid are transmitted to the other one and vice versa with a binaural data link. In this work, this binaural data link is assumed to be ideal which implies neither distortion nor a delay of the transmitted signals. A multi-channel signal processing can be applied to suppress the noise and interefering signals and thus produce an estimate of the desired signal contained in the microphones. Correspondingly, the objective of the multi-channel filter $\mathbf{w}$ is to suppress the noise and interfering signals as much as possible without distorting the desired speech signal.

1.3 Outline of the thesis

In this thesis, we study two recent multi-channel noise reduction algorithms which are the MWF and SP and discuss their relation to the MVDR beamformer.

The objective of this thesis is to evaluate the MWF and SP algorithms in the context of hearing aids. They are compared and evaluated for speech being corrupted by non-stationary noise. These objectives are met by theoretical (Chapter 2) and experimental (Chapters 3 and 4) evaluations.

The thesis is organized as follows: In **Chapter 2**, the two recent multi-channel noise reduction algorithms, MWF and SP are evaluated theoretically and their relation to the MVDR beamformer will be discussed. First, the MVDR beamformer is formulated in the frequency domain, later, the MWF is also formulated in the frequency domain and then it will be shown that the MWF can be decomposed into an MVDR beamformer and a single-channel post filter. Finally, the SP algorithm is also formulated in the frequency domain and it will be shown that SP is an MVDR beamformer when only noise reduction is desired and no dereverberation is considered.

Chapter 3 describes the experimental setup for the simulations and the performance measures used to evaluate the multi-channel noise reduction algorithms.

In **Chapter 4**, experimental results are discussed for the multi-channel algorithms MWF and SP when applied to binaural hearing aids.

In **Chapter 5**, the thesis concludes with a summary and an outlook for future work.

Chapter 2

Noise Reduction by Multi-channel Wiener Filtering and Spatial Prediction

Multi-channel noise reduction algorithms can achieve a high noise reduction performance and therefore, the hearing aids are fitted with multiple microphones. Several multi-channel noise reduction algorithms have been proposed in literature. An interesting solution is provided by the MVDR beamformer which minimizes the output power of the beamformer under a single linear constraint that the desired speech signal arriving from the assumed direction is processed without any distortion [4].

Another multi-channel algorithm is the MWF which produces a minimum mean squared error (MMSE) estimate of the desired speech component contained in a reference microphone signal. Using the MWF some speech distortion is inevitably obtained. The MWF can be slightly modified to provide an explicit tradeoff between speech distortion and noise reduction, resulting in the so-called Speech Distortion Weighted Multi-channel Wiener Filter (SDW-MWF). It can be shown that the MWF can be further decomposed into an MVDR beamformer and a single-channel Wiener post filter [5].

The idea behind the SP algorithm is that the speech signals received by the microphones come from a unique source, so the target signal received by the other microphones can be predicted from a reference microphone signal. The optimal spatial prediction vector can be found in the MMSE sense [7]. Exploiting the spatial prediction vector, a multi-channel noise reduction algorithm can be designed that minimizes target signal distortion (similar to the MVDR beamformer).

Both, the MWF and SP methods utilize the speech and noise PSD matrices to estimate the desired speech component in the reference microphone signal. The main difference between these two algorithms is that the MWF estimates the desired speech component in the MMSE sense and some distortion is inevitable with this algoritm, whereas, in the SP method, the signals are preprocessed by a spatial prediction vector which can be found in the MMSE sense and because of this preprocessing, the speech distortion error can be imposed to be zero.

In this chapter, the two recent multi-channel noise reduction algorithms, MWF and SP are evaluated theoretically and their relation to the MVDR beamformer is discussed. This chapter is organized as follows: In **Section 2.1**, we introduce the problem of noise reduction and the notation used throughout the thesis. In **Section 2.2**, the MVDR beamformer is formulated in the frequency domain and we derive the MVDR beamformer when only noise reduction is desired. In **Section 2.3**, the MWF is formulated in the frequency domain and its relation to the MVDR beamformer is discussed. It will be shown that the MWF can be further decomposed into an MVDR beamformer and a single-channel Wiener post filter. In **Section 2.4**, we formulate the SP algorithm in the frequency domain and prove that SP is an MVDR beamformer when only noise reduction is desired (no dereverberation). In **Section 2.5**, the theoretical results from this chapter are summarized.

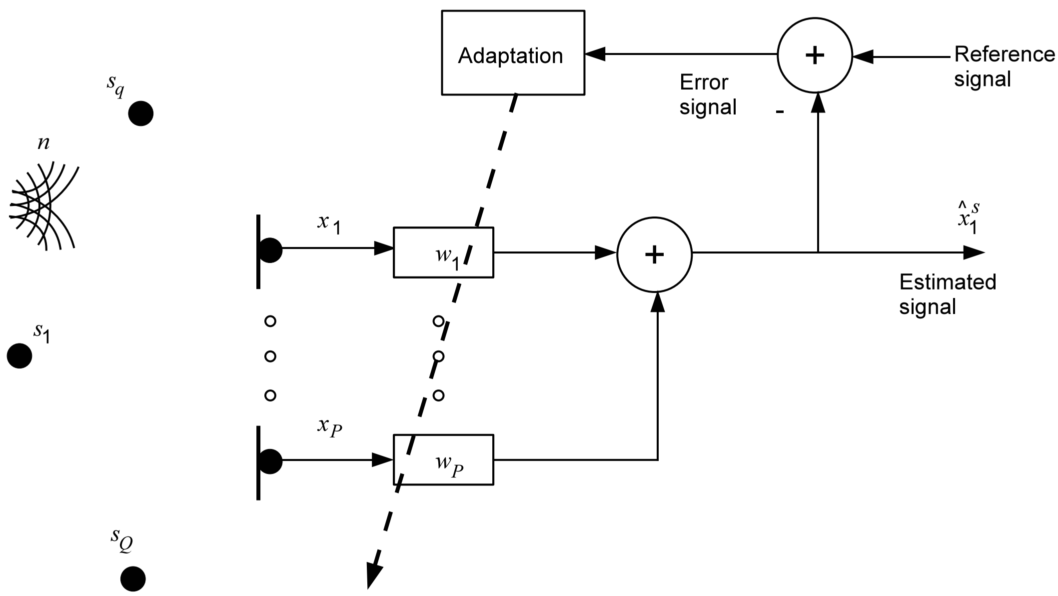


Figure 2.1: Multi-channel noise reduction scheme

2.1 Problem formulation

Consider a multi-channel noise reduction scheme as illustrated in Fig. 2.1. Let P be the number of channels (microphones). Let $s_1(k)$ be the desired source in the presence of interfering signals denoted by $s_2(k)$,...,$s_Q(k)$ and background noise $n(k)$. The received microphone signals in the time domain can be expressed as

$$x_p(k) = \sum_{q=1}^{Q} g_{qp}(k) * s_q(k) + n_p(k), \qquad p = 1, 2, ..., P, \tag{2.1}$$

where k denotes the discrete-time index, $g_{qp}(k)$ is the impulse response from the source $s_q(k)$ to the p-th microphone, * denotes convolution, $n_p(k)$ is the noise at microphone p, and $x_p(k)$ denotes the received signal at the p-th microphone. (2.1) can be rewritten as,

$$
\begin{aligned}
x_p(k) &= \underbrace{g_{1p}(k) * s_1(k)}_{\text{desired speech component}} + \underbrace{\sum_{q=2}^{Q} g_{qp}(k) * s_q(k)}_{\text{interference components}} + \underbrace{n_p(k)}_{\text{background noise}} \\
&= x_p^s(k) + x_p^n(k),
\end{aligned} \tag{2.2}
$$

where $x_p^s(k)$ represents the desired speech component at p-th microphone given by

$$
x_p^s(k) = g_{1p}(k) * s_1(k), \tag{2.3}
$$

and $x_p^n(k)$ represents all interference and noise components at p-th microphone given by

$$
x_p^n(k) = \sum_{q=2}^{Q} g_{qp}(k) * s_q(k) + n_p(k). \tag{2.4}
$$

In the frequency domain, (2.1) can be expressed as,

$$
\begin{aligned}
X_p(w) &= \sum_{q=1}^{Q} G_{qp}(w) S_q(w) + N_p(w) \qquad p = 1, 2, ..., P \\
&= X_p^s(w) + X_p^n(w),
\end{aligned} \tag{2.5}
$$

where $X_p(w)$, $G_{qp}(w)$, $S_q(w)$, $N_p(w)$, $X_p^s(w)$, and $X_p^n(w)$ are the discrete time fourier transforms (DTFTs) of $x_p(k)$, g_{qp}, $s_q(k)$, $n_p(k)$, x_p^s, and x_p^n, respectively. For conciseness, let us omit the frequency variable w from now on. The signal model (2.5) can be summarized in a vector/matrix notation as

$$
\begin{aligned}
\mathbf{x} &= \mathbf{G}^T \mathbf{s} + \mathbf{n} \\
&= \mathbf{x}^s + \mathbf{x}^n,
\end{aligned} \tag{2.6}
$$

where

$\mathbf{x} = [X_1, X_2, ..., X_P]^T$,

$\mathbf{s} = [S_1, S_2, ..., S_Q]^T$,

$\mathbf{n} = [N_1, N_2, ..., N_P]^T$,

$\mathbf{x}^s = [X_1^s, X_2^s, ..., X_P^s]^T$,

$\mathbf{x}^n = [X_1^n, X_2^n, ..., X_P^n]^T$

where superscript $[\cdot]^T$ denotes transpose of a vector or a matrix, and the mixing matrix $\mathbf{G}$ is given by

$$\mathbf{G} = \begin{pmatrix} G_{11} & G_{12} & \cdots & G_{1P} \\ G_{21} & G_{22} & \cdots & G_{2P} \\ \vdots & \vdots & \ddots & \vdots \\ G_{Q1} & G_{Q2} & \cdots & G_{QP} \end{pmatrix}.$$

Without loss of generality, let us consider the first microphone ($p = 1$) as the reference microphone. The microphone signals are filtered by a filter $\mathbf{w}$ to produce an output signal $\hat{X}_1^s$, which represents an estimate of the desired speech component in the reference microphone given by [5],

$$\begin{aligned} \hat{X}_1^s &= \mathbf{w}^H \mathbf{x} \\ &= \mathbf{w}^H \mathbf{x}^s + \mathbf{w}^H \mathbf{x}^n \\ &= \mathbf{w}^H \mathbf{g}_1 S_1 + \mathbf{w}^H \mathbf{x}^n, \end{aligned} \tag{2.7}$$

where
$\mathbf{w} = [W_1, W_2, ..., W_P]^T$,
$\mathbf{g}_1 = [G_{11}, G_{12}, ..., G_{1P}]^T$,
and superscript $[\cdot]^H$ denotes the hermitian of a vector or a matrix.
The error signal e is defined by the difference between the output and desired signal and is given by [3],[4],

$$e = \hat{X}_1^s - cS_1, \tag{2.8}$$

where cS_1 denotes a parameterized desired signal and c is a complex scaling factor which defines the nature of the desired signal. With $c = G_{11}^d$, where G_{11}^d denotes the direct path of impulse response G_{11}, the objective is to suppress noise and interferences as well as to achieve dereverberation [3],[4]. With $c = G_{11}$, the objective is to suppress only noise and interferences (no dereverberation) or in other words, the objective is to recover the desired target speech components in the reference microphone signal $X_1 = G_{11}S_1$ [3],[4].

Using (2.7) and (2.8), the error signal can be written as

$$\begin{aligned} e &= \mathbf{w}^H \mathbf{g}_1 S_1 + \mathbf{w}^H \mathbf{x}^n - cS_1 \\ &= \underbrace{(\mathbf{w}^H \mathbf{g}_1 - c)S_1}_{e_s} + \underbrace{\mathbf{w}^H \mathbf{x}^n}_{e_n}. \end{aligned} \tag{2.9}$$

The error signal can be split up into two parts, e_s, which denotes the error between the true components and the estimated components, and e_n, which denotes the residual noise at the output. The objective of any multi-channel noise reduction algorithm is to minimize the total error denoted by e as well as possible.

2.2 MVDR beamformer

Let us now formulate the MVDR beamformer in the frequency domain. Consider the error signal e defined in (2.9). The mean squared error is given by [3],[4]

$$\begin{aligned} J(\mathbf{w}) &= E[|e|^2] \\ &= E[|e_s^2|] + E[|e_n^2] \\ &= E[|(\mathbf{w}^H \mathbf{g}_1 - c)S_1|^2] + E[|\mathbf{w}^H \mathbf{x}^n|^2] \\ &= |\mathbf{w}^H \mathbf{g}_1 - c|^2 \phi_{s_1 s_1} + \mathbf{w}^H \Phi_{\mathbf{nn}} \mathbf{w}, \end{aligned} \tag{2.10}$$

where

$$\phi_{s_1 s_1} = E[S_1 S_1^*] \tag{2.11}$$

is the power spectral density (PSD) of the signal S_1. $E[\cdot]$ denotes the expectation operator, and

$$\Phi_{\mathbf{nn}} = E[\mathbf{x}^n (\mathbf{x}^n)^H] \tag{2.12}$$

is the PSD matrix of the noise. The objective of the MVDR beamformer is to minimize the noise power at the beamformer output while the desired signal is processed without any distortion. The MVDR beamformer can be formulated by [3],[4]

$$\mathbf{w}_{\mathrm{MVDR}} = \underset{\mathbf{w}}{\operatorname{argmin}} \;\; \mathbf{w}^H \Phi_{\mathbf{nn}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{g}_1 - c = 0. \tag{2.13}$$

The above constrained problem can be converted to an equivalent unconstrained problem, by introducing the Lagrange multiplier λ [4]

$$L(\mathbf{w}, \lambda) = \mathbf{w}^H \Phi_{\mathbf{nn}} \mathbf{w} + \lambda[\mathbf{w}^H \mathbf{g}_1 - c]. \tag{2.14}$$

The MVDR beamformer can be obtained by setting the derivative of (2.14) with respect to $\mathbf{w}$ to zero [3],[4]

$$\mathbf{w}_{\text{MVDR}} = c^* \frac{\Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1}{\mathbf{g}_1^H \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1}. \tag{2.15}$$

where supersript $[\cdot]^*$ denotes complex conjugation. By multiplying the numerator and denominator of (2.15) with $\phi_{s_1 s_1}$ and replacing $\mathbf{g}_1^H \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1$ with $\text{tr}(\Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1^H \mathbf{g}_1)$, we get [3],[4]

$$\mathbf{w}_{\text{MVDR}} = c^* \frac{\phi_{s_1 s_1} \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1}{\phi_{s_1 s_1} \text{tr}(\Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1^H \mathbf{g}_1)}, \tag{2.16}$$

where $\text{tr}(\cdot)$ denotes the trace of a matrix. (2.16) can be further simplified to [3],[4]

$$\mathbf{w}_{\text{MVDR}} = \frac{c^*}{G_{11}^*} \frac{\Phi_{\mathbf{nn}}^{-1} \Phi_{\mathbf{ss}} \mathbf{u}}{\text{tr}(\phi_{\mathbf{nn}}^{-1} \Phi_{\mathbf{ss}})}, \qquad \text{where} \;\; (\phi_{s_1 s_1} \mathbf{g}_1 = \frac{1}{G_{11}^*} \Phi_{\mathbf{ss}} \mathbf{u}), \tag{2.17}$$

where

$$\Phi_{\mathbf{ss}} = E[\mathbf{x}^s (\mathbf{x}^s)^H], \tag{2.18}$$

and

$$\mathbf{u} = [1, 0, ..., 0]^T$$

is a vector of length P. By considering the MVDR beamformer as expressed in (2.17), the beamformer depends only on a single acoustic transfer function G_{11} [3],[4]. When only noise reduction is desired and no dereverberation at all, the MVDR beamformer can be written as [3],[4]

$$\mathbf{w}_{\text{MVDR}} = \frac{\Phi_{\mathbf{nn}}^{-1} \Phi_{\mathbf{ss}} \mathbf{u}}{\text{tr}(\phi_{\mathbf{nn}}^{-1} \Phi_{\mathbf{ss}})}, \tag{2.19}$$

Under the assumption that $\mathbf{x}^s$ and $\mathbf{x}^n$ are mutually uncorrelated, $\Phi_{\mathbf{ss}}$ can be written as

$$\Phi_{\mathbf{ss}} = \Phi_{\mathbf{xx}} - \Phi_{\mathbf{nn}}, \tag{2.20}$$

where $\Phi_{\mathbf{xx}}$ is given by

$$\Phi_{\mathbf{xx}} = E[\mathbf{x}\mathbf{x}^H]. \tag{2.21}$$

By substituting (2.20) in (2.19) and on further simplification, we get [3],[4]

$$\mathbf{w}_{\text{MVDR}} = \frac{\Phi_{\mathbf{nn}}^{-1}\Phi_{\mathbf{xx}} - \mathbf{I}}{\text{tr}(\Phi_{\mathbf{nn}}^{-1}\Phi_{\mathbf{xx}}) - P}. \tag{2.22}$$

where $\mathbf{I}$ is an identity matrix of size $P \times P$, and

2.3 Multi-channel Wiener filtering

Consider the mean squared error $J(\mathbf{w})$ derived in (2.10). The objective of the MWF is to provide an MMSE estimate of either the clean speech source signal (noise reduction and dereverberation) or the speech component in the reference microphone signal (only noise reduction). The MWF is obtained by minimizing the MSE [5],[6]

$$\mathbf{w}_{\text{MWF}} = \underset{\mathbf{w}}{\text{argmin}} \;\; J(\mathbf{w}). \tag{2.23}$$

The solution is obtained by setting the derivative of the cost function $J(\mathbf{w})$ with respect to $\mathbf{w}$ to zero [5],[6]

$$\mathbf{w}_{\text{MWF}} = c^* \Phi_{\mathbf{xx}}^{-1} \phi_{s_1 s_1} \mathbf{g_1}, \tag{2.24}$$

If $c = G_{11}$, i.e., when only noise reduction is desired, then

$$\mathbf{w}_{\text{MWF}} = \Phi_{\mathbf{xx}}^{-1}\Phi_{\mathbf{ss}}\mathbf{u}, \tag{2.25}$$

From (2.20), $\Phi_{\mathbf{ss}}$ can be replaced with $(\Phi_{\mathbf{xx}} - \Phi_{\mathbf{nn}})$, therefore, (2.25) can be further simplified to

$$\begin{aligned} \mathbf{w}_{\text{MWF}} &= \Phi_{\mathbf{xx}}^{-1}(\Phi_{\mathbf{xx}} - \Phi_{\mathbf{nn}})\mathbf{u} \\ &= (\mathbf{I} - \Phi_{\mathbf{xx}}^{-1}\Phi_{\mathbf{nn}})\mathbf{u}. \end{aligned} \tag{2.26}$$

From (2.26), we can say that, when only noise and interference suppression is desired, then no information about the transfer function $\mathbf{g}_1$ is necessary.

2.3.1 Relation to the MVDR beamformer

In this section, the relation between the MWF and the MVDR beamformer is formulated. We show that the MWF can be decomposed into an MVDR beamformer and a single-channel Wiener post filter.

Consider the MWF given by (2.24), which can be rewritten as [5],[6]

$$\begin{aligned}\mathbf{w}_{\mathrm{MWF}} &= c^* \Phi_{\mathbf{xx}}^{-1} \phi_{s_1 s_1} \mathbf{g_1} \\ &= c^* [\Phi_{\mathbf{ss}} + \Phi_{\mathbf{nn}}]^{-1} \phi_{s_1 s_1} \mathbf{g_1} \\ &= c^* [\phi_{s_1 s_1} \mathbf{g_1} \mathbf{g_1}^H + \Phi_{\mathbf{nn}}]^{-1} \phi_{s_1 s_1} \mathbf{g_1}. \end{aligned} \tag{2.27}$$

The MWF can now be factorized into an MVDR beamformer and a single-channel Wiener filter by applying the matrix inversion lemma [5]

$$[\mathbf{A}^{-1} + \mathbf{B}\mathbf{C}^{-1}\mathbf{B}^H]^{-1} = \mathbf{A} - \mathbf{A}\mathbf{B}(\mathbf{C} + \mathbf{B}^H\mathbf{A}\mathbf{B})^{-1}\mathbf{B}^H\mathbf{A}. \tag{2.28}$$

Substituting $\mathbf{A} = \Phi_{\mathbf{nn}}{}^{-1}$, $\mathbf{B} = \sqrt{\phi_{s_1 s_1}}\mathbf{g_1}$, and $\mathbf{C} = 1$ into equation (2.28), the MWF can be transformed into [5]

$$\begin{aligned}\mathbf{w}_{\mathrm{MWF}} &= c^* \left[\Phi_{\mathbf{nn}}{}^{-1} - \frac{\phi_{s_1 s_1} \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1} \mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1}}{1 + \phi_{s_1 s_1} \mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1}} \right] \phi_{s_1 s_1} \mathbf{g_1} \\ &= c^* \left[1 - \frac{\phi_{s_1 s_1} \mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1}}{1 + \phi_{s_1 s_1} \mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1}} \right] \phi_{s_1 s_1} \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1} \\ &= c^* \left[\frac{\phi_{s_1 s_1}}{1 + \phi_{s_1 s_1} \mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1}} \right] \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1} \\ &= \left[\frac{\phi_{s_1 s_1}}{\phi_{s_1 s_1} + (\mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1})^{-1}} \right] \frac{c^* \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1}}{\mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1}}. \end{aligned} \tag{2.29}$$

By taking into account that $(\mathbf{g_1}^H \Phi_{\mathbf{nn}}{}^{-1} \mathbf{g_1})$ is scalar and real valued [5], (2.29) shows that the MWF can be factorized into MVDR beamformer and a real valued scalar factor. Let us analyze the first term of the multi-channel Wiener filter in (2.29). The power spectral density of the desired speech component at the output of the MVDR beamformer is given by [5]

$$
\begin{aligned}
\phi_{s_o s_o} &= \phi_{s_1 s_1} \mathbf{w}_{\mathrm{MVDR}}^H \mathbf{g}_1 \mathbf{g}_1^H \mathbf{w}_{\mathrm{MVDR}} \\
&= \phi_{s_1 s_1} \left| \frac{\mathbf{g}_1^H \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1}{\mathbf{g}_1^H \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1} \right|^2 \\
&= \phi_{s_1 s_1},
\end{aligned} \tag{2.30}
$$

It is shown in (2.30) that the MVDR beamformer gives a distortionless magnitude response [5]. The PSD of the noise component at the output of the beamformer is given by [5]

$$
\begin{aligned}
\phi_{n_o n_o} &= \mathbf{w}_{\mathrm{MVDR}}^H \Phi_{nn} \mathbf{w}_{\mathrm{MVDR}} \\
&= \frac{\mathbf{g}_1^H \Phi_{\mathbf{nn}} \mathbf{g}_1}{(\mathbf{g}_1^H \Phi_{\mathbf{nn}} \mathbf{g}_1)^2} \\
&= \frac{1}{\mathbf{g}_1^H \Phi_{\mathbf{nn}} \mathbf{g}_1},
\end{aligned} \tag{2.31}
$$

Using (2.30) and (2.31), the first term on the righthand side of equation (2.29) can be rewritten as [5]

$$
\frac{\phi_{s_1 s_1}}{\phi_{s_1 s_1} + (\mathbf{g_1}^H {\Phi_{\mathbf{nn}}}^{-1} \mathbf{g_1})^{-1}} = \frac{\phi_{s_1 s_1}}{\phi_{s_1 s_1} + \phi_{n_o n_o}}, \tag{2.32}
$$

which is a single-channel Wiener filter applied at the output of the MVDR beamformer. Therefore, it was proven that the MWF can be factorized into a product of MVDR beamformer and a single-channel Wiener postfilter applied to the beamformer output [5],[6]

$$
\mathbf{w}_{\mathrm{MWF}} = \underbrace{\left[\frac{\phi_{s_1 s_1}}{\phi_{s_1 s_1} + \phi_{n_o n_o}}\right]}_{\text{Wiener postfilter}} \cdot \underbrace{\frac{c^* {\Phi_{nn}}^{-1} \mathbf{g_1}}{\mathbf{g_1}^H {\Phi_{\mathbf{nn}}}^{-1} \mathbf{g_1}}}_{\text{MVDR beamformer}} . \tag{2.33}
$$

2.4 Spatial prediction

Let us now derive the spatial prediction filter. This method exploits the spatial correlations between the speech components to find the optimal filter. The main principle is that the desired speech signal received by the microphones come from a unique source, so the target speech components received by the other microphones can be predicted from a certain reference microphone signal by a spatial prediction vector. By pre-processing the signals by the spatial prediction vector, the speech distortion constraint can be imposed to be zero.

Let us consider that the P speech components can be related to the reference speech signal by [7]

$$\begin{aligned}\mathbf{x}^s &= \mathbf{g}_1 S_1 \\ &= \mathbf{g}_{\mathrm{SP}} X_1^s,\end{aligned} \tag{2.34}$$

where

$$X_1^s = G_{11} S_1, \tag{2.35}$$

and

$$\begin{aligned}\mathbf{g}_{\mathrm{SP}} &= \left[1, \frac{G_{12}}{G_{11}}, ..., \frac{G_{1P}}{G_{11}}\right]^T \\ &= \frac{1}{G_{11}} \mathbf{g}_1.\end{aligned} \tag{2.36}$$

where $\mathbf{g}_{\mathrm{SP}}$ is the spatial prediction vector. The spatial prediction vector can be found in the MMSE sense [7], i.e.,

$$\mathbf{g}_{\mathrm{SP}} = \underset{\mathbf{g}_{\mathrm{SP}}}{\operatorname{argmin}} \; E[(\mathbf{x}^s - \mathbf{g}_{\mathrm{SP}} X_1^s)^H (\mathbf{x}^s - \mathbf{g}_{\mathrm{SP}} X_1^s)], \tag{2.37}$$

The solution is given by [7]

$$\mathbf{g}_{\mathrm{SP}} = \frac{1}{\mathbf{u}^H \Phi_{\mathbf{ss}} \mathbf{u}} \Phi_{\mathbf{ss}} \mathbf{u}. \tag{2.38}$$

The error signal of the multi-channel filter (2.8, 2.9) can now be reformulated with respect to $\mathbf{g}_{\mathrm{SP}}$

$$\begin{aligned}e &= \hat{X}_1^s - c S_1 \\ &= \mathbf{w}^H \mathbf{g}_1 S_1 + \mathbf{w}^H \mathbf{x}^n - c S_1 \\ &= \mathbf{w}^H \mathbf{g}_{\mathrm{SP}} G_{11} S_1 + \mathbf{w}^H \mathbf{x}^n - c S_1 \\ &= (\mathbf{w}^H \mathbf{g}_{\mathrm{SP}} G_{11} - c) S_1 + \mathbf{w}^H \mathbf{x}^n,\end{aligned} \tag{2.39}$$

If $c = G_{11}$, then the error signal e can be written as:

$$\begin{aligned} e &= (\mathbf{w}^H \mathbf{g}_{\mathrm{SP}} - 1) G_{11} S_1 + \mathbf{w}^H \mathbf{x}^n \\ &= (\mathbf{w}^H \mathbf{g}_{\mathrm{SP}} - 1) X_1^s + \mathbf{w}^H \mathbf{x}^n. \end{aligned} \tag{2.40}$$

The objective of spatial prediction filter can be described as [7]

$$\mathbf{w}_{\mathrm{SP}} = \underset{\mathbf{w}}{\operatorname{argmin}} \ \mathbf{w}^H \Phi_{\mathbf{nn}} \mathbf{w} \quad \text{subject to} \quad \mathbf{w}^H \mathbf{g}_{\mathrm{SP}} - 1 = 0. \tag{2.41}$$

which means, in the spatial prediction method, we can impose the speech distortion to be zero. Again, by introducing the lagrange multiplier λ, the equivalent unconstrained problem is given by

$$L(\mathbf{w}, \lambda) = \mathbf{w}^H \Phi_{\mathbf{nn}} + \lambda (\mathbf{w}^H \mathbf{g}_{\mathrm{SP}} - 1). \tag{2.42}$$

The optimal filter can be obtained by setting the derivative of (2.42) with respect to $\mathbf{w}$ to zero [7]

$$\mathbf{w}_{\mathrm{SP}} = \frac{\Phi_{\mathbf{nn}}^{-1} \mathbf{g}_{\mathrm{SP}}}{\mathbf{g}_{\mathrm{SP}}^H \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_{\mathrm{SP}}}. \tag{2.43}$$

2.4.1 Relation to the MVDR beamformer

In this section, we show that spatial prediction is equal to an MVDR beamformer when only noise reduction is considered (no dereverberation). By substituting $\mathbf{g}_{\mathrm{SP}} = \frac{1}{G_{11}} \mathbf{g}_1$ in (2.43), it can be rewritten as

$$\begin{aligned} \mathbf{w}_{\mathrm{SP}} &= \frac{\frac{1}{G_{11}} \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1}{\frac{1}{G_{11} G_{11}^*} \mathbf{g}_1^H \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1} \\ &= G_{11}^* \frac{\Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1}{\mathbf{g}_1^H \Phi_{\mathbf{nn}}^{-1} \mathbf{g}_1}. \end{aligned} \tag{2.44}$$

By comparing (2.44) and (2.15), it can be proven that the spatial prediction principle as formulated above results in an MVDR beamformer for noise and interference suppression only (no dereverberation).

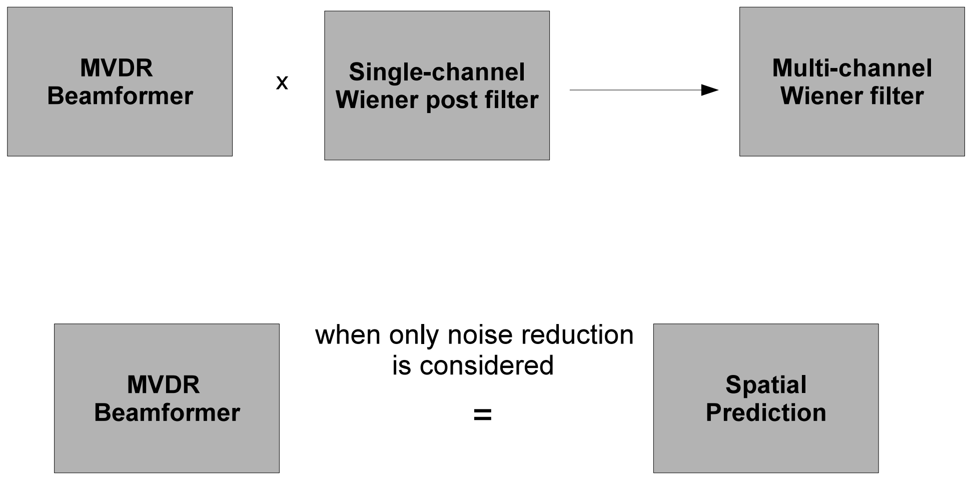


Figure 2.2: Relation of an MVDR beamformer to a multi-channel Wiener filter and Spatial Prediction

2.5 Summary

The theoretical results derived in this chapter are summarized in Fig. 2.2. From (2.33), the MWF is equal to an MVDR beamformer followed by a single-channel Wiener post filter. The single-channel post filter minimizes the squared error of the residual speech and noise components at the output of the MVDR beamformer and therefore, some speech distortion is inevitable. The more the noise is reduced, the more the speech is distorted.

From (2.15) and (2.44), spatial prediction is an MVDR beamfomer when only noise reduction is considered and no dereverberation is taken into account. Since the signals are pre-processed by the spatial prediction vector, the speech distortion error constraint can be imposed to be zero.

Theoretically, the MVDR beamformer and the MWF yield the same SNR, if the input is a narrowband signal [5]. But in scenarios, where speech signals are considered (which are broadband), the post filter may significantly improve the

SNR [5] and therefore, in applications where some amount of speech distortion can be tolerated, the MWF seems to give a better solution. But in the case of hearing aids, where the objective is to maintain the speech intelligibility, the SP algorithm seems to be beneficial (from a theoretical point of view) because of the additional constraint imposed on the speech distortion error to be zero.

Chapter 3

Experimental Setup

In this chapter, the experimental setup for the simulations to evaluate the MWF and SP algorithms, in application to hearing aids is described. This chapter is organized as follows: In **Section 3.1**, the setup for the simulations and the acoustic environments in which the simulations were performed is described. In **Section 3.2**, the implementation of the polyphase filterbank [8] is briefly described. In **Section 3.3**, the estimation of the statistical properties is described and in **Section 3.4**, the performance measures used to assess the performance of the MWF and SP algorithms are defined.

3.1 Setup

We consider two, behind-the-ear (BTE) hearing aids, where each hearing aid (left and right) is equipped with two microphones giving a total number of $P = 4$ microphones. We assume that we have access to all the available microphone signals which implies an ideal binaural data link between both hearing aids to exchange all microphone signals. We apply the multi-channel signal processing to both hearing aids to generate an estimate of the desired speech component for each hearing aid, the right front and the left front microphones were chosen to be the reference microphones for each side.

Simulations were performed for four different acoustic environments. These are an anechoic chamber, a low reverberation chamber (LRC) with a reverberation time $T_{60} \approx 50$ ms, and two living room-like environments namely, MMR and MMRCO with $T_{60} \approx 250$ ms and $T_{60} \approx 400$ ms, respectively. The head-related impulse responses (HRIRs) for the anechoic chamber were taken from the database of Oldenburg University, Germany [9] and the HRIRs for the reverberant environments were taken at the University of Erlangen, Germany (see [10] for more information). The microphone signals were generated by convolving the speech and interfering noise signals of duration 10 seconds with the HRIRs corresponding to their angles of arrival and then summed up together.

All the signals are sampled at a sampling frequency of 16 kHz. Throughout this work it is assumed that all noise components contained in the individual microphones are optimally known, which in practice have to be estimated. The entire signal processing was performed by implementing a polyphase filterbank [8] which is discussed briefly in the following section.

3.2 Polyphase filterbank

3.2.1 Settings for the filterbank

Simulations were performed for a prototype filter length of $L = 128$ and the Discrete Fourier Transform (DFT) length was chosen to be $M = L/2$. The maximum downsampling rate was chosen to be $R = M/4$.

3.2.2 Analysis filterbank

The input signal is divided into consecutive blocks of length L. Then each data block of length L is multiplied by the coefficients of the prototype low-pass filter $h[k]$. The polyphase components of the weighted data block of length L are added and then an M-point inverse discrete fourier transform (IDFT) is applied

to generate the M subband signals by the analysis filterbank.
Each subband can then be modified by multiplying with spectral weights.

3.2.3 Synthesis filterbank

The M polyphase components are obtained by applying an M-point IDFT to the modified subband signals and then the M polyphase components are expanded to a data block of length L by the synthesis filterbank. The synthesized output signal is obtained by multiplying the data block with the synthesis filter coefficients $g[k]$. The reconstructed output signal can be obtained by an overlap-add procedure. The design of prototype lowpass filters (analysis and synthesis) for perfect reconstruction can be found in [8].
For a detailed description on the applied polyphase filterbank please refer to [8].

3.3 Estimation of statistical properties

In order to estimate the PSD matrices necessary for realizing the MWF as well as SP, in each data block instantaneous estimates are calculated as follows

$$\rho_{\mathbf{xx}_{(\kappa)}} = \mathbf{x}_{(\kappa)}(\mathbf{x}_{(\kappa)})^H, \tag{3.1}$$

$$\rho_{\mathbf{nn}_{(\kappa)}} = \mathbf{x}^n{}_{(\kappa)}(\mathbf{x}^n_{(\kappa)})^H, \tag{3.2}$$

where κ represents the frequency subband index. An averaged PSD estimate is then obtained by a first-order recursion

$$\Phi_{\mathbf{xx}_{(\mu,\kappa)}} = \lambda\Phi_{\mathbf{xx}_{(\mu-1,\kappa)}} + (1-\lambda)\rho_{\mathbf{xx}_{(\kappa)}}, \tag{3.3}$$

$$\Phi_{\mathbf{nn}_{(\mu,\kappa)}} = \lambda\Phi_{\mathbf{nn}_{(\mu-1,\kappa)}} + (1-\lambda)\rho_{\mathbf{nn}_{(\kappa)}}, \tag{3.4}$$

where μ represents the frame index and λ is a smoothing parameter chosen close to 1. The value of λ was chosen to be 0.9999 for the filter length $L = 128$.

The speech PSD matrix $\Phi_{\mathbf{ss}}$ is then estimated by (based on the assumption that we have access to all noise components and that the speech and noise are mutually uncorrelated)

$$\Phi_{\mathbf{ss}_{(\mu,\kappa)}} = \Phi_{\mathbf{xx}_{(\mu,\kappa)}} - \Phi_{\mathbf{nn}_{(\mu,\kappa)}}. \tag{3.5}$$

The filters MWF and SP are implemented according to (2.26) and (2.43), respectively, using the estimated PSD matrices. The enhanced binaural output signals are reconstructed by the synthesis filterbank.

3.4 Performance measures

To assess the performance of MWF and SP algorithms, the following performance measures were used:
Signal to interference ratio (SIR) gain ($\mathbf{SIR_{gain}}$) which is calculated by

$$\mathrm{SIR}_{\mathrm{gain}} = \mathrm{SIR}_{\mathrm{out}} - \mathrm{SIR}_{\mathrm{in}}, \tag{3.6}$$

where $\mathrm{SIR}_{\mathrm{in}}$ and $\mathrm{SIR}_{\mathrm{out}}$ are calculated as follows:

$$\mathrm{SIR}_{\mathrm{in}} = \frac{\mathrm{SIR}_{\mathrm{in,left}} + \mathrm{SIR}_{\mathrm{in,right}}}{2}, \tag{3.7}$$

$$\mathrm{SIR}_{\mathrm{out}} = \frac{\mathrm{SIR}_{\mathrm{out,left}} + \mathrm{SIR}_{\mathrm{out,right}}}{2}. \tag{3.8}$$

$\mathrm{SIR}_{\mathrm{in,left}}$ and $\mathrm{SIR}_{\mathrm{in,right}}$ represent the SIR at the reference microphones of the left and right hearing aids respectively, and $\mathrm{SIR}_{\mathrm{out,left}}$ and $\mathrm{SIR}_{\mathrm{out,right}}$ represents the SIR at the output of the left and right hearing aids, respectively. $\mathrm{SIR}_{\mathrm{in,left}}$ and $\mathrm{SIR}_{\mathrm{out,left}}$ are given by

$$\mathrm{SIR}_{\mathrm{in,left}} = 10\log_{10}\left(\frac{\sigma^2_{s,\mathrm{in}}}{\sigma^2_{n,\mathrm{in}}}\right), \tag{3.9}$$

$$\mathrm{SIR}_{\mathrm{out,left}} = 10\log_{10}\left(\frac{\sigma^2_{s,\mathrm{out}}}{\sigma^2_{n,\mathrm{out}}}\right). \tag{3.10}$$

$\sigma^2_{s,\text{in}}$ and $\sigma^2_{s,\text{out}}$ represent the long-term signal power of the desired signal components at the input and output of the left reference microphone respectively, and $\sigma^2_{n,\text{in}}$, $\sigma^2_{n,\text{out}}$ represents the long-term signal power of all the noise and interference components at the input and output of the left reference microphone, respectively. $\text{SIR}_{\text{in,right}}$ and $\text{SIR}_{\text{out,right}}$ are defined similarly.

Speech intelligibility (SI) weigthed SNR gain ($\mathbf{SISNR_{gain}}$) calculates the speech-intelligibility weighted long-term SNR value which is directly related to the standard speech intelligibility index (SII) [11], which is given by

$$\text{SISNR}_{\text{gain}} = \text{SISNR}_{\text{out}} - \text{SISNR}_{\text{in}}, \tag{3.11}$$

where SISNR_{in} and $\text{SISNR}_{\text{out}}$ are calculated as follows [12]:

$$\text{SISNR}_{\text{in}} = \frac{\text{SISNR}_{\text{in,left}} + \text{SISNR}_{\text{in,right}}}{2}, \tag{3.12}$$

$$\text{SISNR}_{\text{out}} = \frac{\text{SISNR}_{\text{out,left}} + \text{SISNR}_{\text{out,right}}}{2}. \tag{3.13}$$

$\text{SISNR}_{\text{in,left}}$ and $\text{SISNR}_{\text{in,right}}$ represents the SI weighted SNR at reference microphones of the left and right hearing aid, respectively, and $\text{SISNR}_{\text{out,left}}$ and $\text{SISNR}_{\text{out,right}}$ represents the SI weighted SNR at the output of the left and right hearing aid respectively. $\text{SISNR}_{\text{in,left}}$ and $\text{SISNR}_{\text{out,left}}$ are given by [12]

$$\text{SISNR}_{\text{in,left}} = \sum_f w_f \left(10\log_{10}\frac{\text{P}_{ss}(f)_{\text{in}}}{\text{P}_{nn}(f)_{\text{in}}}\right)_{-15}^{+15}, \tag{3.14}$$

$$\text{SISNR}_{\text{out,left}} = \sum_f w_f \left(10\log_{10}\frac{\text{P}_{ss}(f)_{\text{out}}}{\text{P}_{nn}(f)_{\text{out}}}\right)_{-15}^{+15}. \tag{3.15}$$

where w_f represents the SII weigthing factor for frequency f according to [11], $\text{P}_{ss}(f)_{\text{in}}$ and $\text{P}_{nn}(f)_{\text{in}}$ represents the long term power spectral density estimates of speech and noise at the input of the left reference microphone, respectively. $\text{P}_{ss}(f)_{\text{out}}$ and $\text{P}_{nn}(f)_{\text{out}}$ represents the long-term power spectral density estimates of speech and noise at the output of the left reference microphone, respectively. Before applying the weighting factor w_f, the SNR is restricted to values between -15 dB and + 15 dB at frequency f which is represented by the notation $()_{-15}^{+15}$ in (3.14) and (3.15). $\mathrm{SISNR}_{\mathrm{in,right}}$ and $\mathrm{SISNR}_{\mathrm{out,right}}$ are defined similarly.

Cepstral distortion (CD) is a measure of distortion based on the cepstrum given by

$$\mathrm{CD} \quad = \frac{\mathrm{CD}_{left} + \mathrm{CD}_{right}}{2}, \tag{3.16}$$

where CD_{left} and CD_{right} denotes the cepstral distortion in dB at the reference microphones of the left and right hearing aid respectively. CD_{left} is given by [13]

$$\mathrm{CD}_{left} = \frac{1}{T}\sum_{\tau=1}^{T} B\sqrt{\sum 2(d_{cep}(\tau) - s_{cep}(\tau))^2}. \tag{3.17}$$

where B is a constant which transforms the measure into dB given by $B = \frac{20}{\log 10}$ [13], T is the frame length, $d_{cep}(\tau)$ is the cepstrum of the estimated output signal at the left reference microphone in the frame τ and $s_{cep}(\tau)$ is the cepstrum of the desired speech signal at the input of the left reference microphone. CD_{right} is defined similarly.

Chapter 4

Experimental Results

In this chapter, experimental results for the MWF and SP algorithms are presented.

The evaluated scenarios are depicted in Fig. 4.1. The desired source s_1 was always located in front of the hearing aid user at approximately 0°. For Scenario 1 (Fig. 4.1a), a single interfering source s_2 was located at 45°. For Scenario 2 (Fig. 4.1b), two interfering sources s_2 and s_3 were located at 45° and -45° respectively. For Scenario 3 (Fig. 4.1c), three interfering sources s_2, s_3, and s_4 were located at 45°, 90°, and -45° respectively. For Scenario 4 (Fig. 4.1d), background babble noise was present in addition to three interfering sources s_2, s_3, and s_4 located at 45°, 90°, and -45° respectively. All the source signals were set to approximately equal long-time signal power.

The performance of the MWF and SP is evaluated for the four different scenarios depicted in Fig. 4.1. The results in terms of $\mathrm{SIR}_{\mathrm{gain}}$, $\mathrm{SISNR}_{\mathrm{gain}}$, and CD for the anechoic chamber, LRC, MMR, and MMRCO are shown in Figs. 4.2, 4.3, 4.4, and 4.5, respectively.

From Fig. 4.2, the performance of MWF and SP is almost similar in terms of $\mathrm{SISNR}_{\mathrm{gain}}$. For Scenario 1, the SP algorithm shows an improvement in $\mathrm{SIR}_{\mathrm{gain}}$ of about 5 dB compared to MWF, but it also introduces more speech distortion compared to the MWF which can be seen from Fig. 4.2c. For Scenarios 2-4,

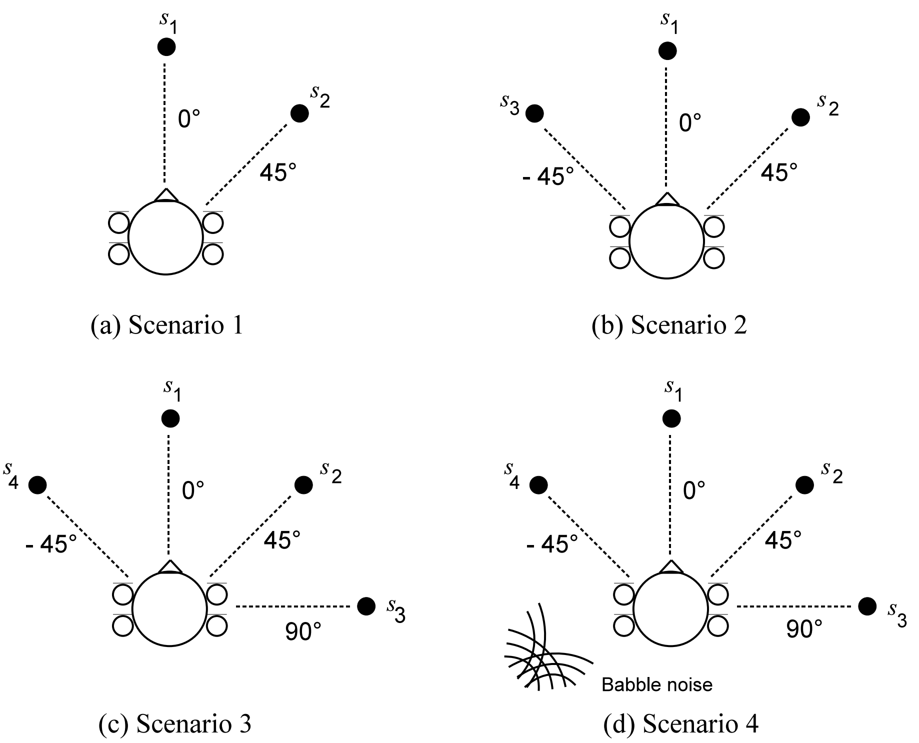
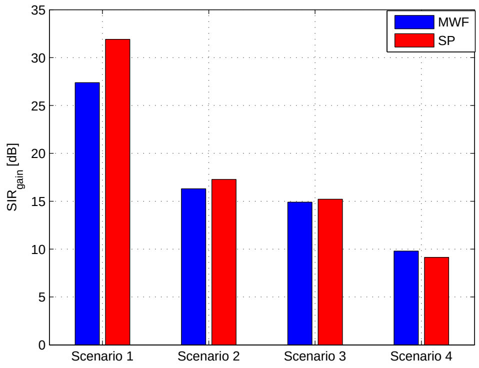


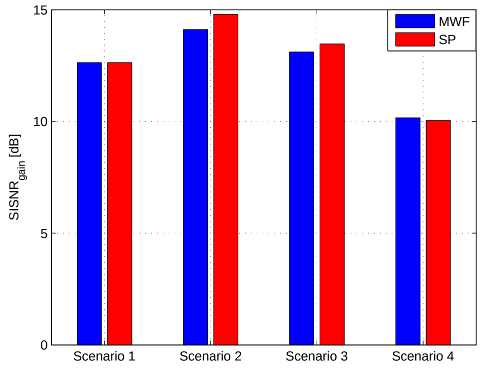
Figure 4.1: Evaluated scenarios

the MWF and SP shows a similar performance in terms of $\mathrm{SIR_{gain}}$. The speech distortion introduced by MWF increases from Scenario 1 to Scenario 4, i.e., from a simple scenario where only one interfering source is present to a more difficult scenario where there are three interfering sources and continuous background noise. The speech distortion introduced by SP is almost constant from Scenario 1 to Scenario 4.

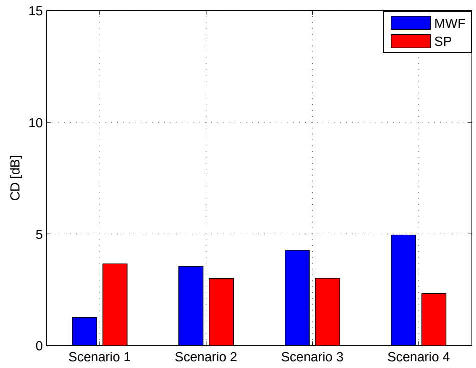
For low reverberant environment LRC, the SP algorithm shows an improvement in terms of $\mathrm{SISNR_{gain}}$ compared to MWF. The performance of MWF and SP is similar in terms of $\mathrm{SIR_{gain}}$ except for the Scenario 1, where the SP attains a gain of about 3 dB more than MWF. But speech distortion introduced by SP is higher for this scenario compared to the MWF which is depicted in Fig. 4.3c. For more dfficult scenarios, for example in Scenario 4, the SP algorithm seems to



(a) SIR gain



(b) SI weighted SNR gain

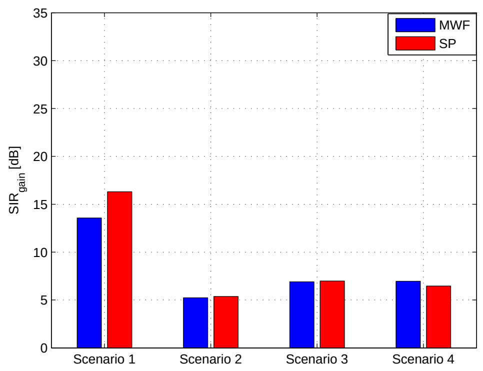


(c) Cepstral distortion

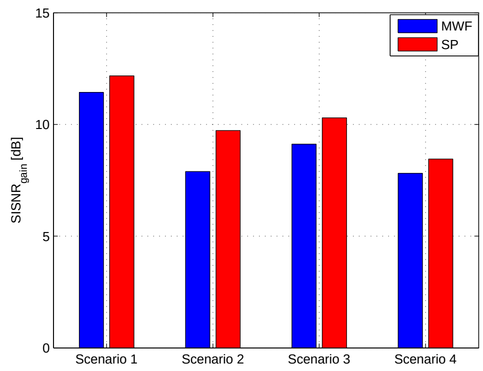
Figure 4.2: Performance of MWF and SP in terms of SIR_{gain}, $\text{SISNR}_{\text{gain}}$, and CD for anechoic chamber with $L = 128$ and $P = 4$

outperform the MWF because it shows an improvement in $\text{SISNR}_{\text{gain}}$ compared to the MWF and also introduces less speech distortion.

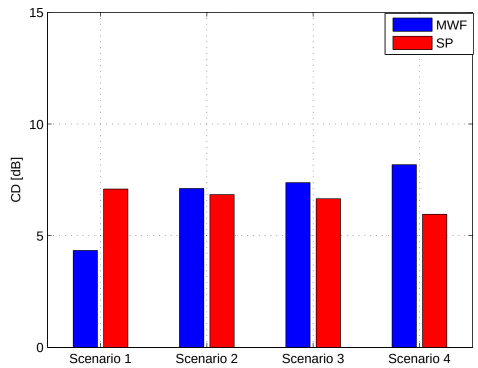
From the anechoic chamber (Fig. 4.2) to the low reverberant environment LRC (Fig. 4.3), one can see a degradation in the performance of both of the algorithms MWF and SP in terms of SIR_{gain} and $\text{SISNR}_{\text{gain}}$ and also speech distortion introduced by both the algorithms is more. For highly reverberated environments MMR and MMRCO, the results are depicted in Figs. 4.4 and 4.5, respectively. The behaviour of the MWF and SP is similar to that of the LRC



(a) SIR gain



(b) SI weighted SNR gain

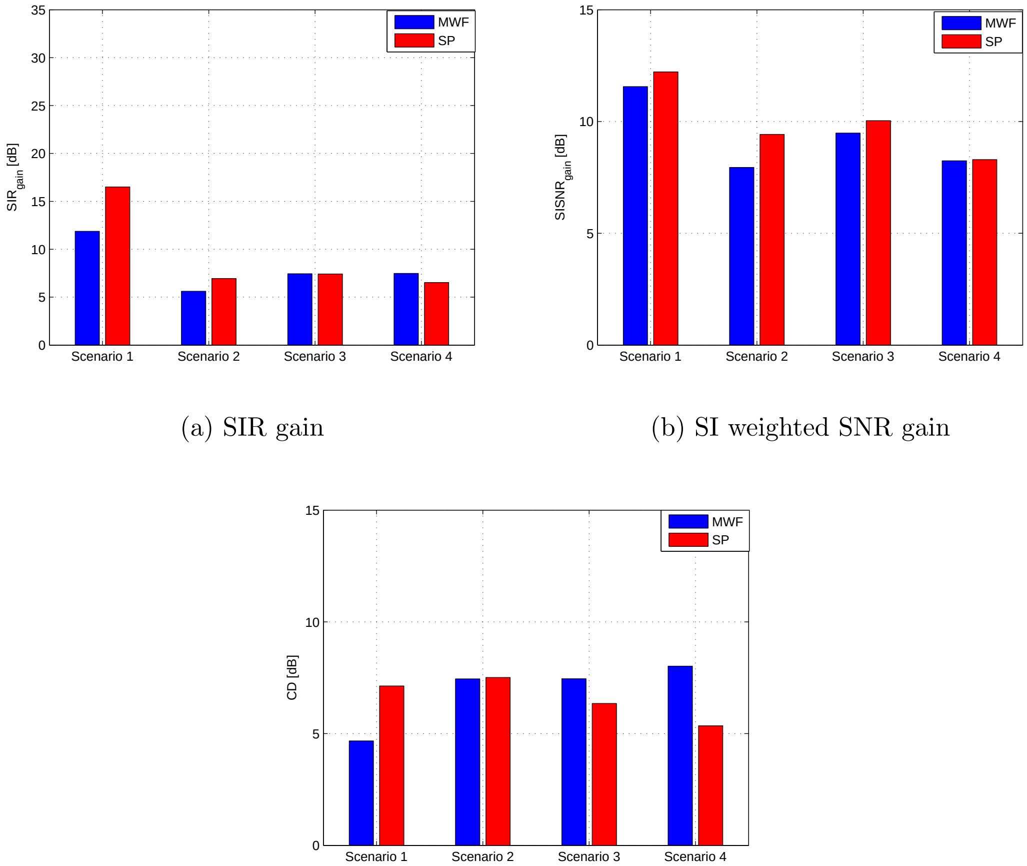


(c) Cepstral distortion

Figure 4.3: Performance of MWF and SP in terms of SIR_{gain}, $\text{SISNR}_{\text{gain}}$, and CD for LRC with $L = 128$ and $P = 4$

except that one can see more degradation in SIR_{gain}, $\text{SISNR}_{\text{gain}}$ and an increase in speech distortion. From these results, the SP algorithm seems to outperform the MWF with an improvement in $\text{SISNR}_{\text{gain}}$ and reduced speech distortion.

Irrespective of the environment, the performance of the MWF degrades from Scenario 1 to Scenario 4. This is due to the fact that the MWF always shows a trade-off between noise reduction and speech distortion and when the interfering sources are close to the target source, some of the target speech components are also suppressed which increases speech distortion and decreases the noise re-



(a) SIR gain

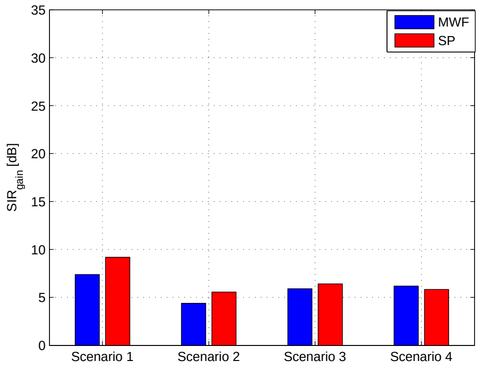
(b) SI weighted SNR gain

(c) Cepstral distortion

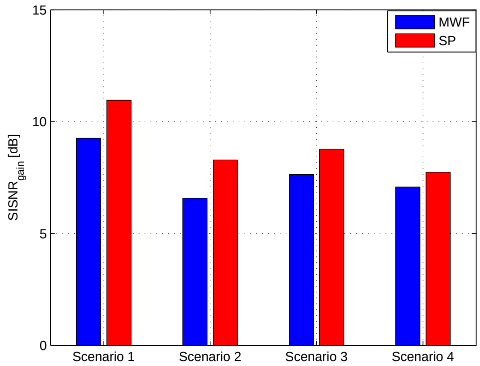
Figure 4.4: Performance of MWF and SP in terms of $\mathrm{SIR}_{\mathrm{gain}}$, $\mathrm{SISNR}_{\mathrm{gain}}$, and CD for MMR with $L = 128$ and $P = 4$

duction performance. This can be seen seen from the entire transfer functions evaluated for the anechoic chamber with filterlength $L = 128$ and $P = 4$ microphones for Scenarios 1-4, depicted in Figs. 4.6-4.9. In these Figures, the x-axis represents the frequency in Hertz and the y-axis represents the possible source positions -180° to +180°.

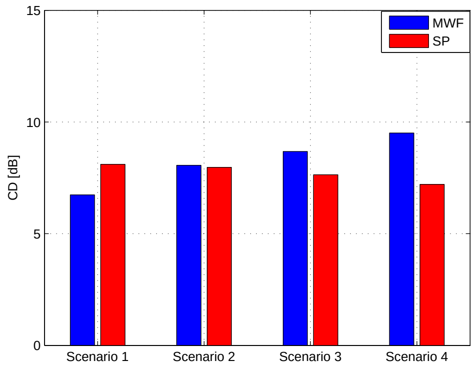
Theoretically, the SP algorithm should not introduce any speech distortion, but from the experimental results, the distortion introduced comes from the errors in the estimation of the spatial prediction vector. Due to the errors in the



(a) SIR gain



(b) SI weighted SNR gain

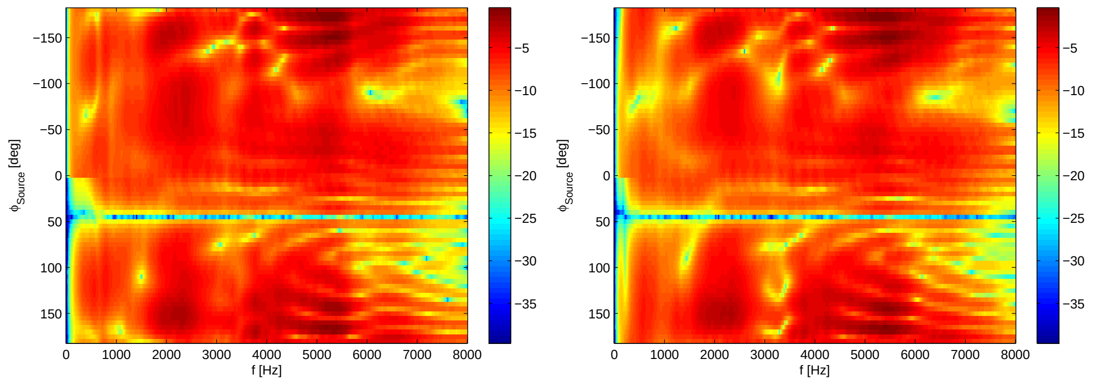


(c) Cepstral distortion

Figure 4.5: Performance of MWF and SP in terms of $\mathrm{SIR}_{\mathrm{gain}}$, $\mathrm{SISNR}_{\mathrm{gain}}$, and CD for MMRCO with $L = 128$ and $P = 4$

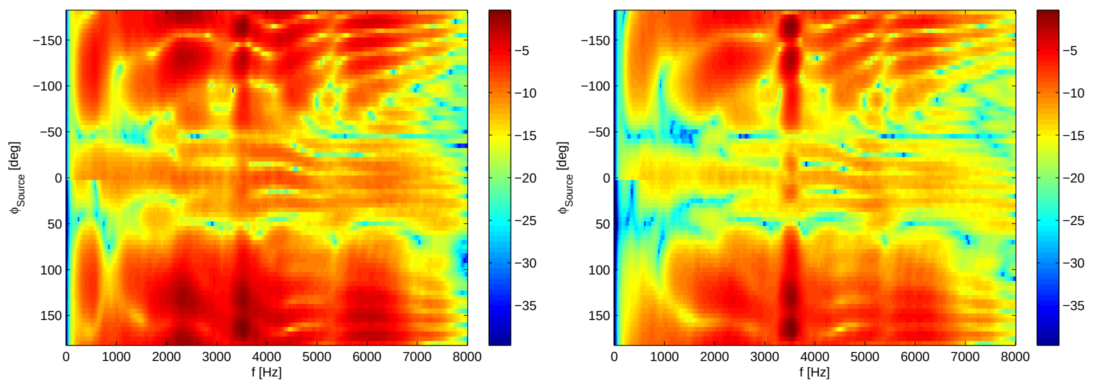
estimation of the spatial prediction vector, the SP algorithm introduces speech distortion and also the noise reduction performance degrades with respect to the scenario. This is verified by repeating the same target signal components contained in a reference microphone signal in all the other microphone signals, i.e., now there is no need to estimate the spatial prediction vector. Figs. 4.10-4.12 shows the obtained results of the SP algorithm with and without the estimation of spatial prediction vector for the LRC, MMR, and MMRCO rooms, respectively.

When there is no need to estimate the spatial prediction vector, the speech dis-



(a) MWF (b) SP

Figure 4.6: Entire transfer function for anechoic chamber with $L = 128$ and $P = 4$ for Scenario 1

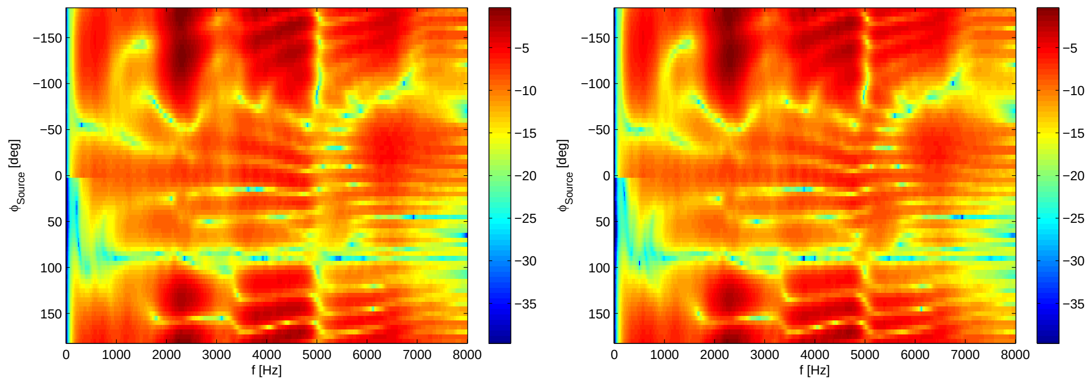


(a) MWF (b) SP

Figure 4.7: Entire transfer function for anechoic chamber with $L = 128$ and $P = 4$ for Scenario 2

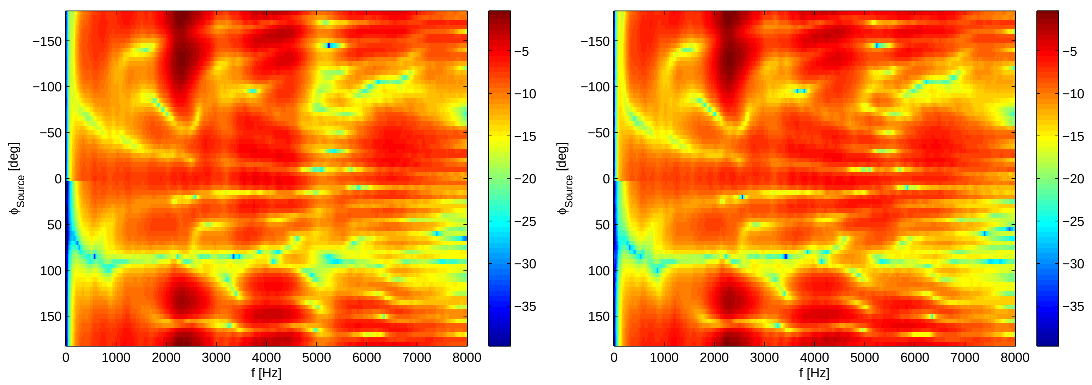
tortion introduced by the SP algorithm is almost negligible and also the $\mathrm{SIR}_{\mathrm{gain}}$ is improved for Scenario 2 and 3. The $\mathrm{SIR}_{\mathrm{gain}}$ degrades for Scenario 4 because of the diffuse babble noise.

The results in terms of $\mathrm{SIR}_{\mathrm{gain}}$, $\mathrm{SISNR}_{\mathrm{gain}}$, and CD for the anechoic chamber with total number of microphones $P = 3$ (both hearing aids are equipped with



(a) MWF (b) SP

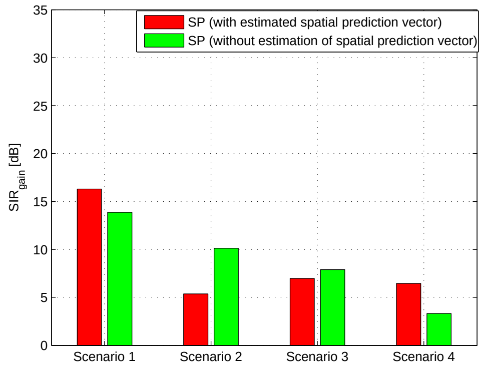
Figure 4.8: Entire transfer function for anechoic chamber with $L = 128$ and $P = 4$ for Scenario 3



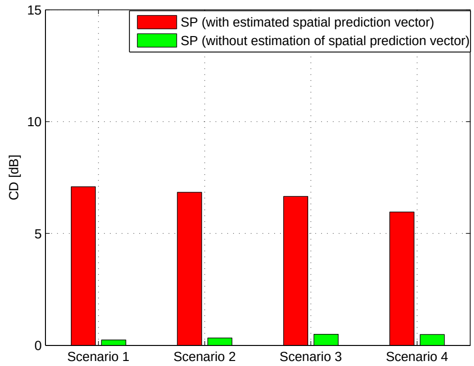
(a) MWF (b) SP

Figure 4.9: Entire transfer function for anechoic chamber with $L = 128$ and $P = 4$ for Scenario 4

two microphones but we transmit only one signal from the right hearing aid to the left hearing aid and vice versa) and $P = 2$ (each hearing aid equipped with one microphone and the signal from the right hearing aid is transmitted to the left hearing aid and vice versa) are shown in Figs. 4.13 and 4.14, respectively. The behaviour of MWF and SP is similar to that discussed when the total number of

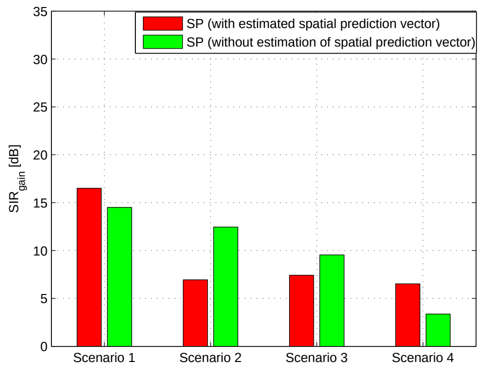


(a) SIR gain

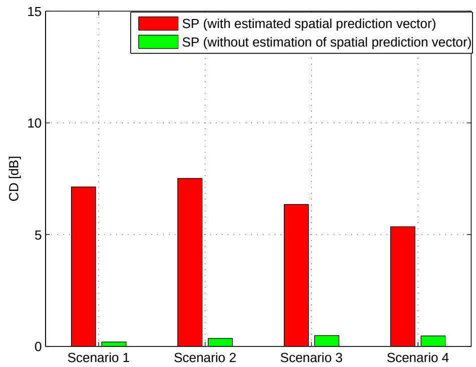


(b) Cepstral distortion

Figure 4.10: Performance of SP (with and without estimation of spatial prediction vector) in terms of $\mathrm{SIR}_{\mathrm{gain}}$ and CD for LRC with $L = 128$ and $P = 4$



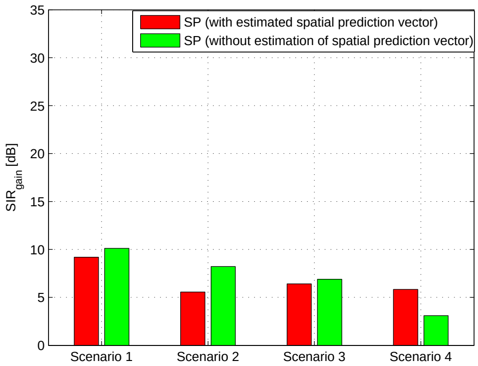
(a) SIR gain



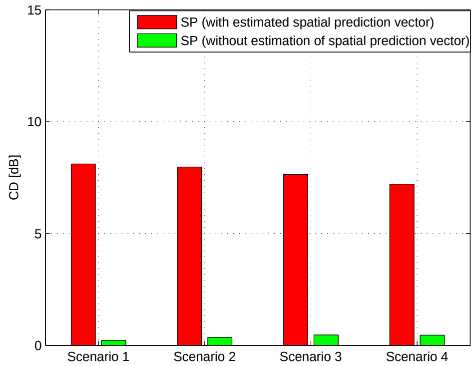
(b) Cepstral distortion

Figure 4.11: Performance of SP (with and without estimation of spatial prediction vector) in terms of $\mathrm{SIR}_{\mathrm{gain}}$ and CD for MMR with $L = 128$ and $P = 4$

microphones $P = 4$. The performance of the MWF and SP in terms of $\mathrm{SIR}_{\mathrm{gain}}$ is similar. The speech distortion introduced by the MWF increases from Scenario 1 to Scenario 4 and the speech distortion introduced by the SP algorithm is almost constant from Scenario 1 to Scenario 4. By comparing the results shown in Figs. 4.2, 4.13, and 4.14, it can be seen that by using more microphones,



(a) SIR gain



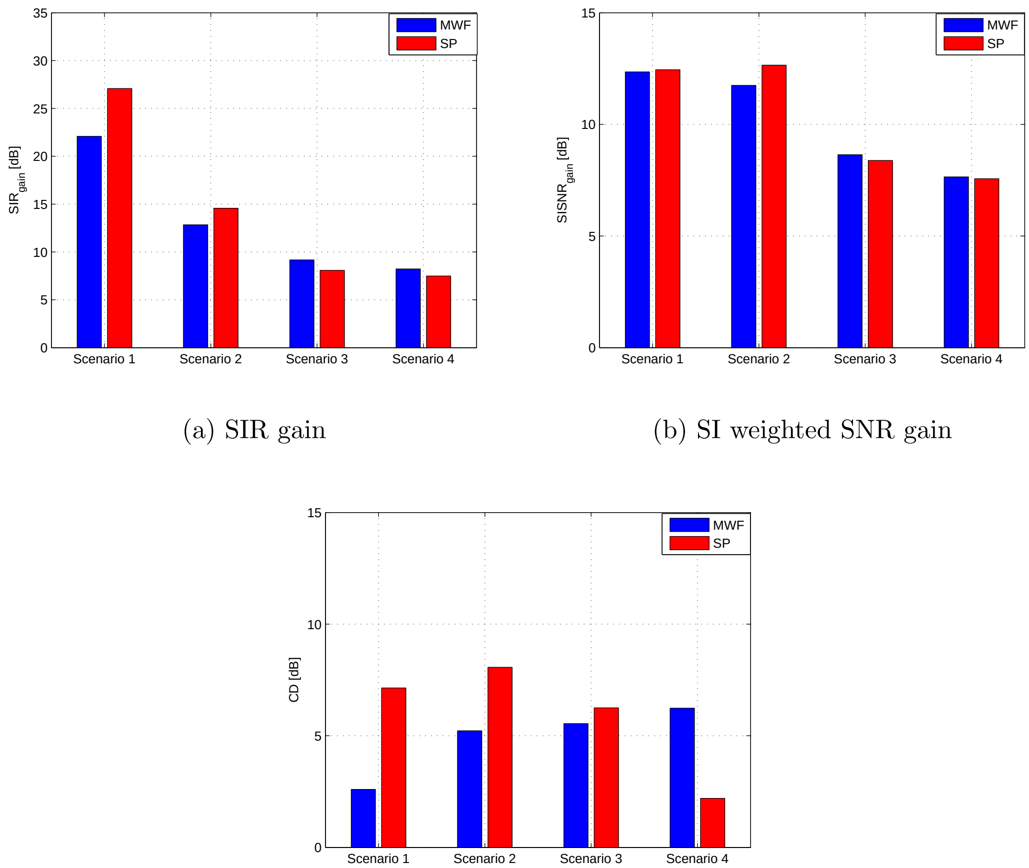
(b) Cepstral distortion

Figure 4.12: Performance of SP (with and without estimation of spatial prediction vector) in terms of SIR_{gain} and CD for MMRCO with $L = 128$ and $P = 4$

SIR_{gain} and $\text{SISNR}_{\text{gain}}$ are improved. The speech distortion also decreases with the increase in number of microphones. Therefore, by using more microphone signals, the noise reduction performance can be improved. However, the speech distortion introduced by the SP algorithm evaluated for $P = 2$ microphones is almost negligible. This is due to fact that the errors in the estimation of the spatial prediction vector decrease.

Eventhough larger filterlengths than $L = 128$ are inappropriate for the hearing aid application, it would be interesting from a theoretical point of view. Let us analyse the effect of increasing L on the noise reduction performance in terms of SIR_{gain} and CD. Figs. 4.15 and 4.16 show the effect of increasing L for an anechoic chamber and the MMR with $P = 4$ microphones considered for Scenario 1.

Both, the MWF and SP show an improvement in the SIR_{gain} as the filter length increases. The performance of the algorithms improve because the scattering, head shadow effects etc., can be better compensated with larger filterlengths L. For the anechoic chamber, the speech distortion introduced by MWF is almost nearly constant for all the filterlengths. But the speech distortion introduced by



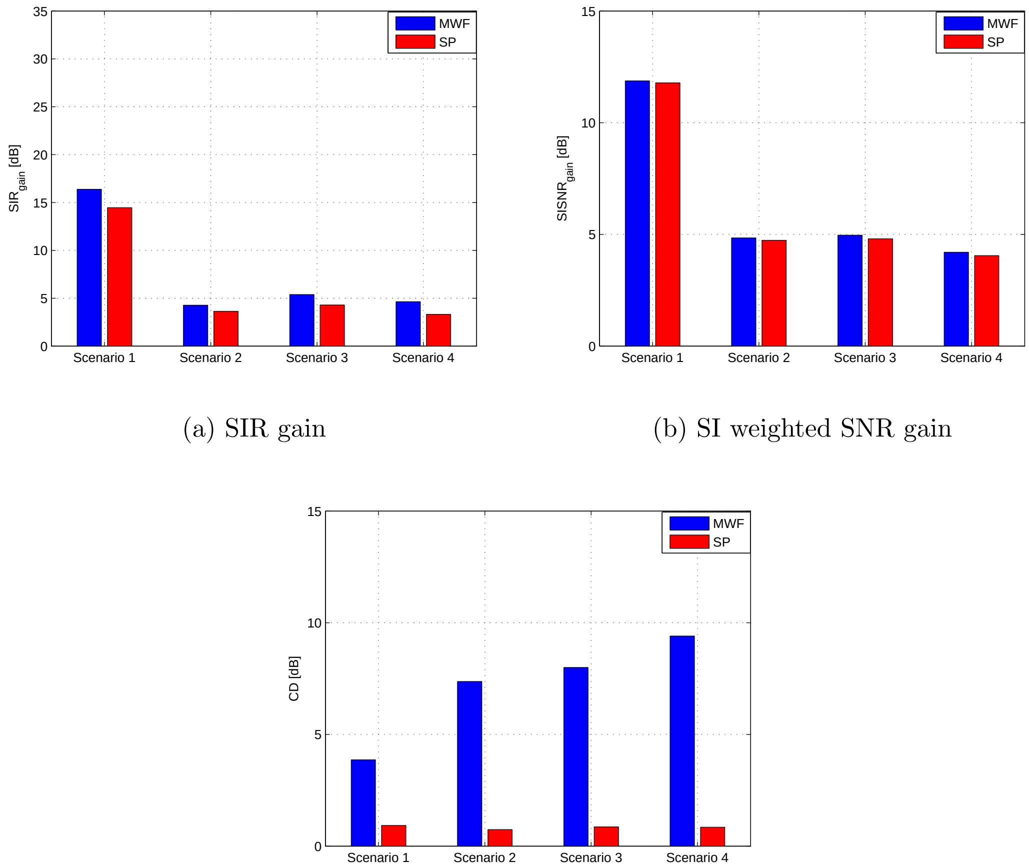
(a) SIR gain

(b) SI weighted SNR gain

(c) Cepstral distortion

Figure 4.13: Performance of MWF and SP in terms of $\mathrm{SIR}_{\mathrm{gain}}$, $\mathrm{SISNR}_{\mathrm{gain}}$, and CD for anechoic chamber with $L = 128$ and $P = 3$

the SP algorithm is more for $L = 128$ and it decreases for $L = 256$ and remains almost nearly constant from then. For the MMR, the speech distortion introduced by both the algorithms decrease as the filterlength increases.

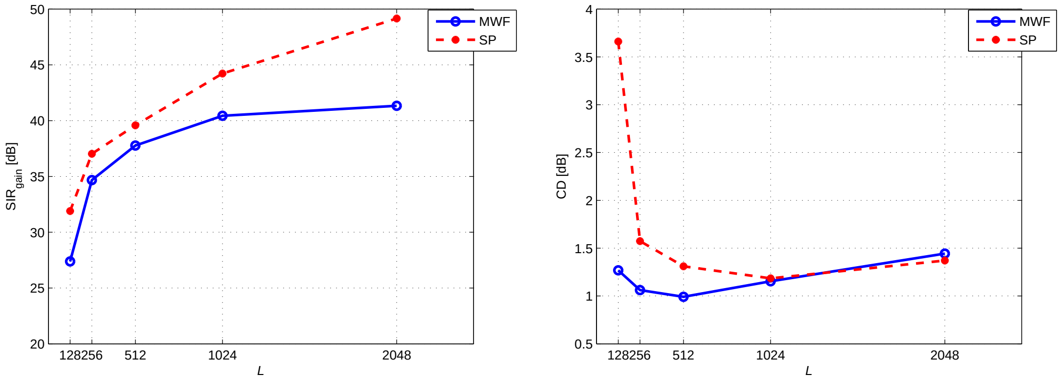


(a) SIR gain

(b) SI weighted SNR gain

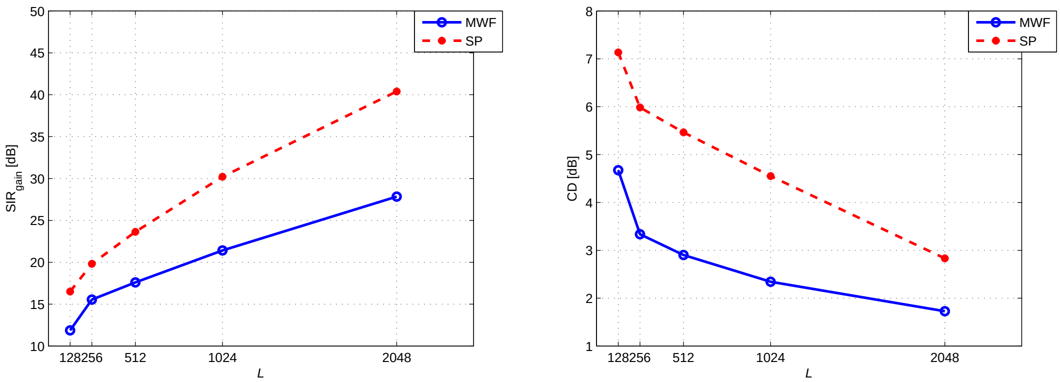
(c) Cepstral distortion

Figure 4.14: Performance of MWF and SP in terms of SIR$_{\text{gain}}$, SISNR$_{\text{gain}}$, and CD for anechoic chamber with $L = 128$ and $P = 2$



(a) SIR gain (b) Cepstral distortion

Figure 4.15: Effect of filterlength L on the performance of MWF and SP in terms of $\mathrm{SIR}_{\mathrm{gain}}$ and CD for anechoic chamber with $P = 4$ considered for Scenario 1



(a) SIR gain (b) Cepstral distortion

Figure 4.16: Effect of filterlength L on the performance of MWF and SP in terms of $\mathrm{SIR}_{\mathrm{gain}}$ and CD for MMR with $P = 4$ considered for Scenario 1

Chapter 5

Conclusions

In this thesis, two recent multi-channel noise reduction algorithms, an MWF and SP are evaluated in the context of binaural hearing aids for speech being corrupted by non-stationary noise.

In the **Chapter 2**, it was shown that the MWF can be decomposed into an MVDR beamformer and a single-channel Wiener post filter. The single-channel post filter minimizes the squared error of the residual speech and noise components at the output of the MVDR beamformer and therefore, some speech distortion is inevitable with this algorithm. It was also proven that SP is an MVDR beamformer when only noise reduction is considered (no dereverberation). Since the signals are pre-processed by a spatial prediction vector, the speech distortion error in the case of the SP can be imposed to be zero. From the theoretical findings in **Chapter 2**, MWF algorithm gives a better solution where some speech distortion is tolerable. Since the speech distortion error can be imposed to be zero in case of the SP algorithm, it seems to give a better solution in applications where the objective is to maintain the speech intelligibilty.

The theoretical findings are verified by experiments conducted on two BTE hearing aids, where an ideal binaural data link between both hearing aids was assumed. The MWF and SP algorithms were evaluated for different scenarios in different environments. From the experimental results in **Chapter 4**, the SP algorithm seems to outperform the MWF by showing an improvement in terms of $\mathrm{SIR}_{\mathrm{gain}}$ and reduced CD for the anechoic chamber as well as for reverberant environments like the LRC, MMR, and MMRCO. From the theoretical findings, the SP algorithm should not introduce any speech distortion, but from the experimental results, the distortion arises from the errors in the estimation of the spatial prediction vector which was verified in **Chapter 4**. In this thesis, it was assumed that all noise components contained in the individual microphones were optimally known, which in practice have to be estimated.

SP algorithm seems to be appealing for the hearing aid applications if it can be realized without any estimation errors for real scenarios. Future work could investigate the problems and errors in the realization and estimation of the SP algorithm.

Bibliography

[1] http://www.hear-it.org, March 2011.

[2] 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 J. Appl. Signal Process.,* vol.2005, pp.2915-2929, 2005.

[3] E. A. P. Habets, J. Benesty, I. Cohen, and S. Gannot, "New Insights Into the MVDR Beamformer in Room Acoustics," *IEEE Transactions on Audio, Speech, and Language Processing, vol.18, no.1, January 2010.*

[4] I. Cohen, J. Benesty, and S. Gannot, *Speech Processing in Modern Communication-Challenges and Perspectives,* Springer-Verlag, Berlin, Germany, 2010.

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

[6] P. Vary, and R. Martin, *Digital Speech Transmission: Enhancement, Coding and Error Concealment*, John Wiley & Sons Ltd, England 2006.

[7] B. Cornelis, M. Moonen, and J. Wouters. "Comparison of Frequency Domain Noise Reduction Strategies Based on Multi-channel Wiener Filtering and Spatial Prediction",*ICASSP, April 2009.*

[8] K. Reindl, Y. Zheng, and W. Kellermann, "Blind Binaural Signal Extraction for Hearing Aids," *Report on project year 1. Technical report, University of Erlangen-Nuremberg, 2010.*

[9] H. Kayser, S. Ewart, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, *Database of multi-channel in-ear and behind-the-ear head-related and Binaural room impulse responses*, University of Oldenburg 2009.

[10] K. Reindl, *Messung von Head-Related Transfer Functions (HRTFs)*, Technical Report, LMS, University of Erlangen-Nuremberg 2008.

[11] *American National Standard Methods for the Calculation of the Speech Intelligibility Index*, American National Standards Institute, New York.

[12] 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 Acoustic Society of America, vol.94, no.5, pp.3009-3010, 1993.*

[13] L. Rabiner and B. Juang, *Fundamentals of Speech Recognition*, Upper Saddle River, NJ: Prentice-Hall, 1993.

[14] H. Dillon, *Hearing aids*, Boomerang Press, 2001.