# Dual Transfer Function Generalized Sidelobe Canceller and Application to Joint Noise Reduction and Echo Cancellation

Submitted in Partial Fulfillment of The Requirements for the Degree of Master of Science in Electrical Engineering

**Gal Reuven**

Submitted to the Senate of the
Technion - Israel Institute of Technology

NISAN, 5766 HAIFA March, 2006

# Acknowledgment

I would like to express my gratitude to all those who gave me the possibility to complete this thesis. First, I want to thank my supervisors. Warm thanks to Dr. Sharon Gannot from the Engineering school in Bar Ilan university, for his interminable patience and dedication, the encouragement to perfection and for stimulating discussions, even in correct Hebrew. My thanks to Dr. Israel Cohen from the department of Electrical Engineering in the Technion, for his assistance and many valuable suggestions along the way.

Thanks to the administrative staff of the Engineering School in Bar Ilan university, and especially to Mrs. Dina Yemini, for creating a pleasant working environment during the last year.

Warm thanks to my good friend, Alon Keren, for his encouragement during long months, where nights and days seemed alike, and for contributing his special skills in Matlab figures.

And finally, above all I would like to give my special thanks to my parents, Rachel and Yitzchak, whom great help and love enabled me to complete this work.

# Contents

# List of Figures

# List of Tables

# Abstract

Man machine interaction requires acoustic interface in order to provide full duplex hands-free communication. In many practical environments a speech signal received by an array of microphones is contaminated by both additive nonstationary interfering signal and stationary noise. The environment is often reverberant, and the received signals are generally corrupted by the room impulse responses. Desired signal quality might significantly deteriorate in the presence of these interferences.

We consider a microphone array located in a reverberant room, where general transfer functions (TFs) relate the source signal and the microphones. The array is used for enhancing a speech signal contaminated by a competing speech signal as well as stationary noise signal. The previously proposed transfer-function generalized sidelobe canceller (TF-GSC) exploits the nonstationarity of the speech signal to enhance it when the undesired interfering signals are stationary. Unfortunately, the TF-GSC is rendered useless when a nonstationary interference is present.

In this work, we propose an adaptive beamformer, based on the TF-GSC, that is suitable for cancelling nonstationary interferences in noisy reverberant environments, denoted Dual TF-GSC (DTF-GSC). We modify two of the TF-GSC components to enable suppression of the nonstationary undesired signal. A modified fixed beamformer is designed to block the nonstationary interfering signal while maintaining the desired speech signal. A modified blocking matrix is designed to block both the desired signal and the nonstationary interference. It is shown that it is sufficient to use the ratio between the different TFs rather

than the TFs themselves in order to implement the suggested algorithm. A novel approach is proposed to estimate the blocking matrix during double talk frames. Experimental results show that the algorithm is working well in various real-life scenarios.

A performance analysis of the DTF-GSC performance is presented as well. A general expression of the enhancer output is first derived. This expression is used to evaluate three figures of merit, i.e., the amount of deviation of the desired signal from its nominal value, the noise reduction and the interference cancellation abilities. The expression depends on the TFs involved, the noise field type and the quality of estimation of the TFs ratios.

In many cases the competing signal is due to the coupling of the loudspeaker and microphone in hands-free communication, i.e., an echo signal. In that case, a reference signal to a direct measurement of the echo signal can be exploited. While echo signals alone can be suppressed successfully by an acoustic echo canceller (AEC) and an adaptive beamformer can reduce the noise, the AEC performance is impaired significantly due to the noise and the adaptive beamformer suffers from the echo signal.

Hence, a TF-GSC based scheme is suggested to jointly reduce noise and cancel echoes. This scheme is evaluated through a series of experiments in single and double talk situations, and compared to two additional cascade schemes, implemented in the frequency domain.

# Glossary

## Acronyms

**AEC** Acoustic Echo cancellation

**ATF** Acoustic Transfer Function

**BF** Beamformer

**BLMS** Block least mean square

**DoA** Direction of Arrival

**DTF-GSC** Dual Transfer Function Generalized Sidelobe Canceller

**ETF-GSC** Echo Transfer Function Generalized Sidelobe Canceller

**GSC** Generalized Sidelobe Canceller

**GSVD** Generalized singular value decomposition

**HOS** Higher Order Statistics

**LCMV** Linearly constrained minimum variance

**LMS** least mean square

**SER** Signal to Echo Ratio

**SNR** Signal to Noise Ratio

**TF** Transfer Function

**TF-GSC** Transfer Function Generalized Sidelobe Canceller

**QRD-LSL** QR decomposition-based least squares lattice

**Abbreviations**

**AEC-BF scheme** Noise and echo reduction cascade scheme where the Acoustic Echo Canceller precede the beamformer

**BF-AEC scheme** Noise and echo reduction cascade scheme the beamformer precede the Acoustic Echo Canceller

**Double talk** Periods in time when both the desired and competing speech signal are active

**Single talk** Periods in time when only the desired speech signal is active

**Symbols**

$*$ Convolution

$(\cdot)^*$ Conjugation

$(\cdot)^\dagger$ Conjugation transpose

$\widehat{(\cdot)}$ Estimated value

$\perp$ Perpendicular spaces

$\overset{\text{FIR}}{\longleftarrow}$ Finite length impulse response operator

$z_m(t)$ The $m$-th microphone signal

$s_1(t)$ The desired speech signal

$s_2(t)$ The nonstationary interfering signal

$e(t)$ The available echo signal measured at the loudspeaker

$n_m(t)$ The stationary noise signal at the $m$-th microphone

$a_m(t)$ The acoustical impulse responses of the $m$-th microphone to the desired speech source

$b_m(t)$ The acoustical impulse responses of the $m$-th microphone to the nonstationary interference (or echo) source

$d_m(t)$ The acoustical impulse responses of the $m$-th microphone to the directional stationary noise source

$Z_m(t, e^{j\omega})$ STFT of the $m$-th microphone signal

$S_1(t, e^{j\omega})$ STFT of the desired speech signal

$S_2(t, e^{j\omega})$ STFT of the nonstationary interfering signal

$E(t, e^{j\omega})$ STFT of the echo signal signal

$N_m(t)$ STFT of the stationary noise signal at the $m$-th microphone

$A_m(e^{j\omega})$ The ATFs from the desired source to the $m$-th microphone

$B_m(e^{j\omega})$ The ATFs from the interference source (or echo) to the $m$-th microphone

$D_m(e^{j\omega})$ The ATFs from the directional stationary noise source to the $m$-th microphone

$\boldsymbol{Z}(t, e^{j\omega})$ Vector formulation of $Z_m(t, e^{j\omega})$ elements

$\boldsymbol{A}(t, e^{j\omega})$ Vector formulation of $A_m(t, e^{j\omega})$ elements

$\boldsymbol{B}(t, e^{j\omega})$ Vector formulation of $B_m(t, e^{j\omega})$ elements

$\boldsymbol{D}(t, e^{j\omega})$ Vector formulation of $D_m(t, e^{j\omega})$ elements

$\boldsymbol{N}(t, e^{j\omega})$ Vector formulation of $N_m(t, e^{j\omega})$ elements

$Y(t, e^{j\omega})$ Beamformer output

$Y_{s_1}(t, e^{j\omega})$ The desired signal part in the beamformer output

$Y_{s_2}(t, e^{j\omega})$ The nonstationary interference signal part in the beamformer output

$Y_n(t, e^{j\omega})$ The stationary noise signal part in the beamformer output

$\Phi_{\boldsymbol{ZZ}}(t, e^{j\omega})$ Power spectral density of the microphone signals

$\mathcal{F}^*(e^{j\omega})$ Arbitrary filter response

$\nabla_{W^*}$ Derivative with respect to $\boldsymbol{W}^*$

$\lambda_i$ The $i$-th Lagrange multiplier

$\mathbf{W}^{opt}(t, e^{j\omega})$ Linearly constrained minimum variance beamformer

$\|\boldsymbol{X}(e^{j\omega})\|_{\Phi}$ Weighted norm of a vector $\boldsymbol{X}(e^{j\omega})$

$\rho_{\Phi}(e^{j\omega})$ The cosine of the angle between the vectors $\boldsymbol{A}(e^{j\omega})$ and $\boldsymbol{B}(e^{j\omega})$ in a weighted inner product space

$P(e^{j\omega})$ The projection matrix to the $\left[ \begin{array}{c|c} \boldsymbol{A}(e^{j\omega}) & \boldsymbol{B}(e^{j\omega}) \end{array} \right]$

$\rho(e^{j\omega})$ The coherence function (or the cosine of the angle between the vectors $\boldsymbol{A}(e^{j\omega})$ and $\boldsymbol{B}(e^{j\omega})$ in an inner product space)

$\mathbf{F}(e^{j\omega})$ Matched beamformer

$\mathcal{N}(e^{j\omega})$ The null space of $\left[ \begin{array}{c|c} \boldsymbol{A}(e^{j\omega}) & \boldsymbol{B}(e^{j\omega}) \end{array} \right]$

$\Lambda(e^{j\omega})$ The constraint hyperplane

$\mathcal{R}(e^{j\omega})$ The column space of $\left[ \begin{array}{c|c} \boldsymbol{A}(e^{j\omega}) & \boldsymbol{B}(e^{j\omega}) \end{array} \right]$

$\boldsymbol{W}_0(t, e^{j\omega})$ Constraint hyperplane component of the beamformer

$\boldsymbol{V}(t, e^{j\omega})$ Noise hyperplane component of the beamformer

$\mathcal{H}(e^{j\omega})$ Blocking matrix

$\boldsymbol{G}(t, e^{j\omega})$ Adaptive noise canceller filters

$Y_{\mathbf{MBF}}(t, e^{j\omega})$ The matched beamformer component of the beamformer output

$Y_{\mathbf{NC}}(t, e^{j\omega})$ The noise canceller component of the beamformer output

$\boldsymbol{U}(t, e^{j\omega})$ The reference noise signals

$Q_m(e^{j\omega})$ $m$-th element of first row of the blocking matrix $\mathcal{H}(e^{j\omega})$

$L_m(e^{j\omega})$ $m$-th element of second row of the blocking matrix $\mathcal{H}(e^{j\omega})$

$\Phi_{\boldsymbol{UU}}$ Power spectral density of the noise reference signals

$\Phi_{\boldsymbol{U}Y}(t, e^{j\omega})$ Cross power spectral density of the noise reference signals and beamformer output signal

$\mu$ NLMS step size

$\eta$ Recursive power estimation forgetting factor

$P_{\mathbf{est}}(t, e^{j\omega})$ Power estimation

$H_m(e^{j\omega})$ Unbiased estimate of $\frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}$

$\Phi_{z_i z_j}(e^{j\omega})$ Cross power spectral density of the $i$-th and the $j$-th microphone signals

$\Phi_{u_m z_1}(e^{j\omega})$ Cross power spectral density of the $m$-th noise reference signal and the first microphone signal

$\mathbf{K}$ The number of frames in the interval

$\varepsilon_m^{(k)}(e^{j\omega})$ The estimation error in the $k$-th frame

$\hat{\Phi}_{z_i z_j}^{(k)}(e^{j\omega})$ The estimation of the $k$-th frame of $\Phi_{z_i z_j}(e^{j\omega})$

$T_{60}$ The reverberation time

$S_1NR$ The signal to noise ratio for the desired signal $s_1(t)$

$S_2NR$ The signal to noise ratio for the nonstationary interference signal $s_2(t)$

$S_1S_2R$ The ratio between the power of $s_1(t)$ and the power of $s_2(t)$

$\Phi_{yy}(t, e^{j\omega})$ Power spectral density of the beamformer output

$\Phi_{\boldsymbol{NN}}(t, e^{j\omega})$ Power spectral density of the noise signals

$\Phi_{\boldsymbol{U}Y_{\mathbf{MBF}}}(t, e^{j\omega})$ Cross power spectral density of the noise reference signals and the beamformer output

$T_1$ Time intervals in which only the desired signal is active

$T_2$ Time intervals in which only the nonstationary interference signal is active

$T_n$ Time intervals in which only the noise signal is active

$\mathbf{DEV}(t, e^{j\omega})$ Deviation of the beamformer output to nominal value

$\theta$ Direction of arrival

$d_{ij}$ Distance between two microphones ($i$,$j$)

$\Gamma_{N_i N_j}(e^{j\omega})$ The cross-coherence function between signals received by two sensors ($i$,$j$) with distance $d_{ij}$

$\Gamma(e^{j\omega})$ The coherence matrix

$\mathbf{NIR}(t, e^{j\omega})$ Normalized interference reduction

$\boldsymbol{G}^N(t, e^{j\omega})$ Adaptive noise canceller filters in the ANEC block

$\boldsymbol{G}^E(t, e^{j\omega})$ Adaptive echo canceller filters in the ANEC block

$\boldsymbol{G}^H(t, e^{j\omega})$ Adaptive echo canceller filters in the BU block

# Chapter 1

# Introduction

## 1.1 Motivation and Background

In many practical environments a speech signal received by an array of microphones is contaminated by both additive nonstationary interfering signal and stationary noise. The environment is often reverberant, and the received signals are generally corrupted by the room impulse responses. Desired signal quality might significantly deteriorate in the presence of these interferences. As a result, the intelligibility of the desired speech signal is impaired, and speech compression and speech recognition systems performance degrades.

The problem of multiple constrains beamforming in a noisy and reverberant environment is addressed in this work. *Linearly constrained minimum variance* (LCMV) [1] beamformer has found numerous applications in the field of speech enhancement in the recent three decades. The most attractive implementation of the LCMV is the generalized sidelobe canceller (GSC) [2]. In this structure the constraint and the minimization are decoupled, yielding a simple but yet powerful tool for handling the problem. A comprehensive survey of beamforming methods can be found in [3]. In most speech enhancement applications the beamformer is constrained to produce a dominant response towards the assumed speech source location, while minimizing the response for all other directions. However, in reverberant environment a single direction of arrival cannot be determined and the desired signal impinges on the array from several directions due to the many

reflections from objects in the room. This problem might be alleviated by using a complex *acoustic transfer function* (ATF) rather than just a simple delay for modelling the propagation of the speech signal in the reverberant room. Affes and Grenier [4] used subspace tracking method for estimating the time invariant ATF. The non-stationarity of the speech signal was exploited by Gannot *et al.* [5] for estimating the transfer function ratio. This extension to the classical GSC structure, nicknamed the *transfer-function generalized sidelobe canceller* (TF-GSC), was used for enhancing speech signal deteriorated by a single *stationary* interference signal in an arbitrary transfer function (TF) enclosure.

In many sound environments the assumption that the interference signal is stationary cannot be met. In these applications interference sources might be a near-end competing speech or a far-end echo signal, rather than a noise signal. The *transfer-function generalized sidelobe canceller* (TF-GSC) proposed in [5] uses beamforming and nonstationarity to enhance speech signal deteriorated by a single stationary interference signal in an arbitrary transfer function (TF) enclosure. When a second, directional, non-stationary interference signal is also present, the algorithm is unable to distinguish between the desired signal and the interference and therefore render useless.

The problem in which both a desired signal and a competing speech signal are received by a microphone array in a noisy and reverberant enclosure can be viewed as a convolutive *blind source separation* (BSS) problem. In this case, the desired and the competing speech signals are filtered by the room impulse response and contaminated by an additive stationary noise sources before being mixed.

Several approaches were taken to handle this problem. Speech signal non-stationarity is exploited by Parra and Spence [6] to obtain a nonlinear minimization problem. Both permutation and gain ambiguity problems, encountered by the frequency domain approach is alleviated by imposing an FIR structure on the mixing filters. When the number of microphones is larger than the number of sources, Parra and Alvino [7] propose to combine beamforming techniques

to circumvent the problem of permutations ambiguities by imposing geometrical constraints on the solution. In the two sources case Kurita *et al.* [8] and Ikram and Morgan [9] use the null location of the beam-pattern to mitigate the permutation problem. Knaak *et al.* [10] incorporate the beamforming constrains into the FastICA algorithm[11], based on *higher order statistics* (HOS).

Our problem is closely related to the echo cancellation problem as well. In these problems a joint effort of mitigating the echo signal and reducing the noise level is required. However, the two tasks generally contradict each other [12], especially in situations where both signals are active (usually, denoted a *double talk situation*). However, it should be stressed that in the echo cancellation problem a separate measurement of the interference signal is available and can be used to improve the performance of the overall system.

Affes and Grenier proposed in [13] a GSC structure for double talk situations. They presented a distortionless fixed beamformer constrained to cancel the echo, and a blocking matrix constrained to block both the desired signal and echo signal. The TFs are estimated using subspace tracking methods. These estimates are used to construct both the fixed beamformer and the blocking matrix.

In [12] two frequency domain schemes for joint echo cancellation and noise reduction are presented. Both contain the TF-GSC beamformer proposed in [5] and a block *least mean square* (LMS) *acoustic echo canceller* (AEC). Following Kellermann [14] , the first scheme comprises multi-channel AEC followed by a beamformer, while the second comprises a beamformer followed by a single channel AEC as a post-filter. A series of simulations using real speech recordings showed that the first scheme outperforms the second one. Two additional schemes for noise reduction and echo cancellation are proposed and compared in [15] by Doclo and Moonen. The first scheme includes a multi-channel AEC followed by a *generalized singular value decomposition* (GSVD) based beamformer. The second scheme incorporates the far-end echo reference into the GSVD beamformer. Simulations indicate that the first scheme outperforms the second one. Rombouts and Moonen [16] combine the speech enhancement and echo cancellation tasks

in one integrated scheme. The optimization problem defined by this scheme is solved adaptively using a QR decomposition-based least squares lattice (QRD-LSL) algorithm. It is shown that the performance of the integrated scheme is superior to the performance of traditional (cascading) schemes, while complexity is kept at an affordable level.

## 1.2 Overview of the Thesis

In this contribution we present a novel method for cancelling two interference signals, one stationary and the second non-stationary, applicable in problems in which the algorithm in [5] fails. Unlike the BSS related methods the proposed algorithm does not suffer from gain and permutation ambiguities. We modify two of the TF-GSC components to enable suppression of the nonstationary undesired signal. A modified fixed beamformer is designed to block the nonstationary interfering signal while maintaining the desired speech signal. A modified blocking matrix is designed to block both the desired signal and the nonstationary interference. We introduce a novel method for updating the blocking matrix in double talk situations, which exploits the nonstationarity of both the desired and interfering speech signals. Experimental results demonstrate the performance of the proposed algorithm in noisy and reverberant environments.

Performance analysis of the DTF-GSC is presented. A general expression of the enhancer output is first derived. The expression depends on the TFs involved, the noise field and the quality of estimation of the TFs ratios. This analytical expression is used to evaluate three figures of merit, i.e. the amount of deviation imposed on the desired signal, the noise reduction and the interference reduction abilities.

Finally, we show how this structure can be easily modified to handle the joint echo suppression and noise reduction tasks, by incorporating the available echo signal into the system. This scheme is evaluated through a series of experiments in single and double talk situations, and compared to two additional

schemes, implemented in the frequency domain. It is shown that the TF-GSC based scheme outperforms significantly the two others while comparing the echo cancellation performance. Similar noise reduction performance are obtained in all three schemes.

The structure of this work is as follows. In Chapter 2 we first formulate the problem of dual-source interference cancelling in a general acoustical transfer function (ATF) environment and suggest a solution. Some experimental results in practical scenarios are presented. Chapter 3 is devoted to analytical evaluation of the expected performance. We discuss the joint echo suppression and noise reduction problem as an application of the TF-GSC beamformer in Chapter 4. Finally, in Chapter 5, we conclude the work and propose subjects for future research.

# Chapter 2

# Dual Source Transfer Function Generalized Sidelobe Canceller

Full duplex hands-free man/machine interface often suffers from directional nonstationary interference, such as a competing speaker, as well as stationary interferences which may comprise both directional and non-directional signals. The transfer-function generalized sidelobe canceller (TF-GSC) exploits the nonstationarity of the speech signal to enhance it when the undesired interfering signals are stationary. Unfortunately, the TF-GSC is rendered useless when a nonstationary interference is present. In this chapter, we propose an adaptive beamformer, based on the TF-GSC, that is suitable for cancelling nonstationary interferences in noisy reverberant environments. We modify two of the TF-GSC components to enable suppression of the nonstationary undesired signal. A modified fixed beamformer is designed to block the nonstationary interfering signal while maintaining the desired speech signal. A modified blocking matrix is designed to block both the desired signal and the nonstationary interference. We introduce a novel method for updating the blocking matrix in double talk situations, which exploits the nonstationarity of both the desired and interfering speech signals. Experimental results demonstrate the performance of the proposed algorithm in noisy and reverberant environments.

## 2.1 Introduction

In many practical environments a speech signal received by an array of microphones is contaminated by both additive nonstationary interfering signal and stationary noise. The environment is often reverberant, and the received signals are generally corrupted by the room impulse responses. Gannot *et al.* [5] assumed that the background noise is stationarity, and proposed to estimate the ATF ratio by exploiting the nonstationarity of the speech signal. Compared with the conventional GSC, the resulting *transfer function generalized sidelobe canceller* (TF-GSC) is of practical utility when enhancing a speech signal deteriorated by *stationary* interfering signals in an arbitrary ATF enclosure. However, when using the TF-GSC algorithm in the presence of a second directional nonstationary interference signal, the overall interference reduction performance severely deteriorates. Theoretically, the MBF steers the desired signal, while the ANC eliminates the interference speech signal component in the MBF output, since the noise reference signals contain interference components. In practical however, the desired signal ATF estimation, based on exploiting the nonstationarity characteristics of the speech signal, impairs due to the second speech signal. In addition, when the noise level is high compared to the interference signal, the ANC cannot eliminate the interference component in the MBF output, since the interference components in the reference signals are masked by the noise signal components.

In this chapter, we present an adaptive beamformer, based on the TF-GSC [5], for cancelling two interference signals, one stationary and the other nonstationary, in a reverberant environment[1]. A closed-form optimal solution is first derived using constrained optimization techniques, and an adaptive recursive algorithm is then developed to enable tracking of time variations in the environment and the ATFs. We modify two of the TF-GSC components, namely the *fixed beamformer* and the *blocking matrix*, to allow the beamformer to sup-

[1] A preliminary version of this work was presented in [17].

press the nonstationary interference. The modified fixed beamformer, denoted matched beamformer (MBF), is designed to block the nonstationary interference while maintaining the desired speech signal. The modified BM blocks both the desired signal and the nonstationary interference. As in the TF-GSC structure, the adaptive noise canceller employs the reference signals generated by the blocking matrix to cancel the residual stationary noise at the MBF output. A novel method for updating the blocking matrix in double talk situations is proposed as well, which exploits the nonstationarity of both the desired and interfering speech signals. Therefore, when using the DTFGSC algorithm rather than the TFGSC, the blocking matrix can be updated during double talk frames. In addition, even if the noise level is stronger than the interference level, the interference can be significantly reduced. Experimental results demonstrate the performance of the proposed algorithm in noisy and reverberant environments. The perceptual quality of the desired speech signal is retained in the enhanced signal, while the stationary and nonstationary interferences are well suppressed (audio sample files are available on-line [18]).

The structure of this chapter is as follows. In Section 2.2, we formulate the problem of a dual-source interference cancelling in a general acoustical transfer function environment. In Section 2.3, we derive a linearly constraint beamformer specifically designed for suppressing undesired interference signals. In Section 2.4 we describe the proposed algorithm. In Section 2.5, we address the problem of ATF estimation, and finally in Section 2.6 we present some experimental results which demonstrate the performance of the proposed algorithm in noisy and reverberant environments.

## 2.2 Problem formulation

Let us consider an array of sensors in a noisy and reverberant environment. We assume that the received signals include three components, a desired speech source, a directional nonstationary interference signal (e.g. competing speech) and a sta-

tionary noise signal, which can be either directional, non-directional or a combination thereof. Our goal is to reconstruct the desired speech signal from received reverberated signals. Let $s_1(t)$ denote the desired speech signal, let $s_2(t)$ denote the nonstationary interfering signal, and let $a_m(t)$ and $b_m(t)$ represent the acoustical impulse responses of the $m$-th microphone to the desired speech source and the nonstationary interference source, respectively. The $m$-th microphone signal is given by

$$z_m(t) = a_m(t) * s_1(t) + b_m(t) * s_2(t) + n_m(t);\ m = 1, \ldots, M \tag{2.1}$$

where $n_m(t)$ is the (directional or nondirectional) stationary noise signal at the $m$-th microphone, and $*$ denotes convolution. The analysis frame duration is chosen such that the signal may be considered stationary over the analysis frame. Typically, the impulse responses $a_m(t)$ and $b_m(t)$ are slowly changing in time and can be considered stationary over the analysis frame.

In the *short time Fourier transform* (STFT) domain, (2.1) can be approximately rewritten [2] as:

$$Z_m(t, e^{j\omega}) \approx A_m(e^{j\omega})S_1(t, e^{j\omega}) + B_m(e^{j\omega})S_2(t, e^{j\omega}) + N_m(t, e^{j\omega}) \\ m = 1, \ldots, M \tag{2.2}$$

where $Z_m(t, e^{j\omega}), S_1(t, e^{j\omega}), S_2(t, e^{j\omega})$ and $N_m(t, e^{j\omega})$ are the STFT of the respective signals. $A_m(e^{j\omega})$ and $B_m(e^{j\omega})$ are the ATFs from the desired source and interference source to the $m$-th microphone, respectively, which are assumed hereinafter time invariant over the observation period. A vector formulation of (2.2) is

$$\boldsymbol{Z}(t, e^{j\omega}) = \boldsymbol{A}(e^{j\omega})S_1(t, e^{j\omega}) + \boldsymbol{B}(e^{j\omega})S_2(t, e^{j\omega}) + \boldsymbol{N}(t, e^{j\omega}) \tag{2.3}$$

[2] The approximation sign in (2.2) can be replaced with equality when the length of the signal is large compared with the length of the filter. This assumption cannot be met in our case, when considering the ATFs themselves. However, this approximation is sufficient for the ATFs estimation procedure. Moreover, while applying the algorithm, the ATFs ratios are used rather than the ATFs themselves. These ratios are considered to be much shorter. It should be noted that each filtering operation is implemented using the overlap and save procedure, avoiding cyclic convolution effect.

where

$$\begin{aligned}
\boldsymbol{Z}(t,e^{j\omega}) &= \left[\, Z_1(t,e^{j\omega})\ Z_2(t,e^{j\omega})\ \cdots\ Z_M(t,e^{j\omega}) \,\right]^T \\
\boldsymbol{A}(e^{j\omega}) &= \left[\, A_1(e^{j\omega})\ A_2(e^{j\omega})\ \cdots\ A_M(e^{j\omega}) \,\right]^T \\
\boldsymbol{B}(e^{j\omega}) &= \left[\, B_1(e^{j\omega})\ B_2(e^{j\omega})\ \cdots\ B_M(e^{j\omega}) \,\right]^T \\
\boldsymbol{N}(t,e^{j\omega}) &= \left[\, N_1(t,e^{j\omega})\ N_2(t,e^{j\omega})\ \cdots\ N_M(t,e^{j\omega}) \,\right]^T .
\end{aligned}$$

Our problem is to reconstruct the desired speech signal $S_1(t,e^{j\omega})$ (or a filtered version thereof) from the noisy observations $\boldsymbol{Z}(t,e^{j\omega})$.

## 2.3 Optimal solution based on constrained optimization

In this section, we derive a linearly constrained beamformer, specifically designed for suppressing undesired interference signals. We first obtain a closed-form linearly constrained minimum variance beamformer, and then derive an adaptive solution. We initially assume that the ATFs are known and in Section 2.5 we derive their estimates based on the nonstationarity of the speech signals.

Let $W_m^*(t,e^{j\omega})\ ; m=1,\ldots,M$ be a set of $M$ filters,

$$\boldsymbol{W}^{\dagger}(t,e^{j\omega}) = \left[\, W_1^*(t,e^{j\omega})\ W_2^*(t,e^{j\omega})\ \cdots\ W_M^*(t,e^{j\omega}) \,\right]$$

where $^*$ denotes conjugation and $^\dagger$ denotes conjugation transpose. A beamformer is realized by filtering each sensor output $Z_m(t,e^{j\omega})$ by $W_m^*(t,e^{j\omega})\ ; m=1,\ldots,M$ and summing the outputs:

$$\begin{aligned}
Y(t,e^{j\omega}) &= \boldsymbol{W}^{\dagger}(t,e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}) = \boldsymbol{W}^{\dagger}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})S_1(t,e^{j\omega}) \\
&+\boldsymbol{W}^{\dagger}(t,e^{j\omega})\boldsymbol{B}(e^{j\omega})S_2(t,e^{j\omega}) + \boldsymbol{W}^{\dagger}(t,e^{j\omega})\boldsymbol{N}(t,e^{j\omega}) \\
&\triangleq Y_{s_1}(t,e^{j\omega}) + Y_{s_2}(t,e^{j\omega}) + Y_n(t,e^{j\omega}) \qquad (2.4)
\end{aligned}$$

where $Y_{s_1}(t,e^{j\omega})$ represents the desired signal part, $Y_{s_2}(t,e^{j\omega})$ is the directional interference part and $Y_n(t,e^{j\omega})$ is the stationary noise part. The output power is

given by:

$$E\{Y(t,e^{j\omega})Y^*(t,e^{j\omega})\} = E\{\boldsymbol{W}^\dagger(t,e^{j\omega})\boldsymbol{Z}(t,e^{j\omega})\boldsymbol{Z}^\dagger(t,e^{j\omega})\boldsymbol{W}(t,e^{j\omega})\}$$
$$= \boldsymbol{W}^\dagger(t,e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{W}(t,e^{j\omega})$$

where $\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega}) = E\{\boldsymbol{Z}(t,e^{j\omega})\boldsymbol{Z}^\dagger(t,e^{j\omega})\}$. We want to minimize the output power subject to the following constraints:

$$Y_{s_1}(t,e^{j\omega}) = \boldsymbol{W}^\dagger(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})S_1(t,e^{j\omega}) = \mathcal{F}^*(e^{j\omega})S_1(t,e^{j\omega})$$
$$Y_{s_2}(t,e^{j\omega}) = \boldsymbol{W}^\dagger(t,e^{j\omega})\boldsymbol{B}(e^{j\omega})S_2(t,e^{j\omega}) = 0 \tag{2.5}$$

where $\mathcal{F}^*(e^{j\omega})$ is an arbitrary filter response. This filter may imply any filtering operation imposed on the output due to reverberation. We thus have the following minimization problem

$$\min_{\boldsymbol{W}} \left\{\boldsymbol{W}^\dagger(t,e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{W}(t,e^{j\omega})\right\} \tag{2.6}$$
$$\text{subject to } \boldsymbol{W}^\dagger(t,e^{j\omega})\boldsymbol{A}(e^{j\omega}) = \mathcal{F}^*(e^{j\omega}) \text{ and } \boldsymbol{W}^\dagger(t,e^{j\omega})\boldsymbol{B}(e^{j\omega}) = 0.$$

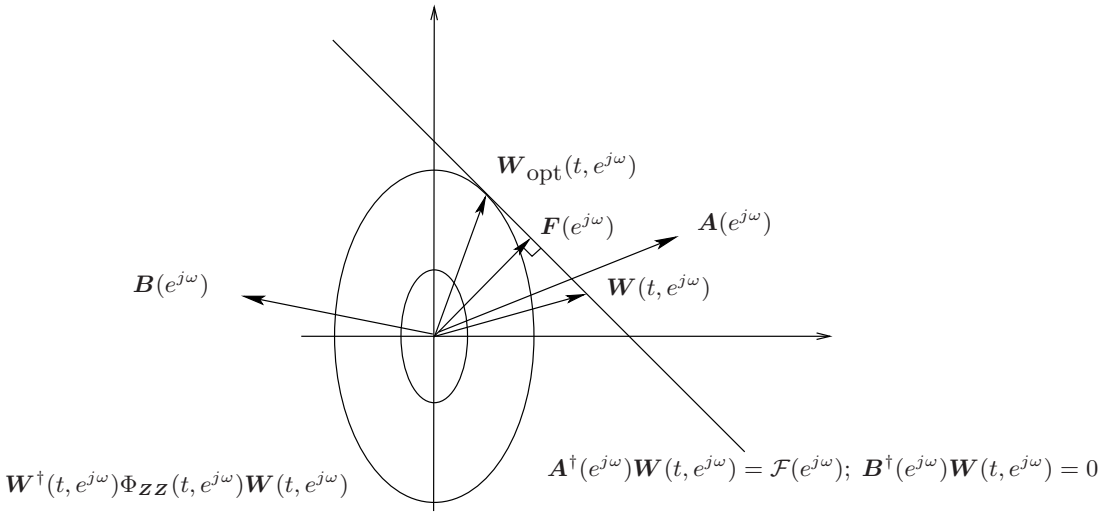


Figure 2.1: Constrained minimization.

The minimization in (2.6) is depicted in Fig. 2.1. The tangent point of the equi-power contours with the constraint line is the optimum vector of beamformer filters. Solution to the constrained minimization problem is obtained by

minimizing the complex Lagrangian:

$$\mathcal{L}(\mathbf{W}) = \mathbf{W}^{\dagger}(t,e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) + \lambda_1\left[\mathbf{W}^{\dagger}(t,e^{j\omega})\mathbf{A}(e^{j\omega}) - \mathcal{F}^{*}(e^{j\omega})\right] \quad (2.7)$$
$$+\lambda_1^{*}\left[\mathbf{A}^{\dagger}(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega})\right] + \lambda_2\mathbf{W}^{\dagger}(t,e^{j\omega})\mathbf{B}(e^{j\omega}) + \lambda_2^{*}\mathbf{B}^{\dagger}(e^{j\omega})\mathbf{W}(t,e^{j\omega}).$$

Setting the derivative with respect to $\boldsymbol{W}^{*}$ to zero (see for instance [19]) we obtain

$$\nabla_{W^{*}}\mathcal{L}(\mathbf{W}) = \Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) + \lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega}) = 0 \quad (2.8)$$

and since $\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})$ is usually invertible[3], $\mathbf{W}(t,e^{j\omega})$ can be written as:

$$\mathbf{W}(t,e^{j\omega}) = -\Phi_{\boldsymbol{ZZ}}^{-1}(t,e^{j\omega})\left[\lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega})\right]. \quad (2.9)$$

Imposing the constraints on $\mathbf{W}(t,e^{j\omega})$ and solving for the Lagrange multipliers yields (see Appendix A)

$$\mathbf{W}^{opt}(t,e^{j\omega}) = \mathcal{F}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}^{-1}(t,e^{j\omega})\frac{\frac{\boldsymbol{A}(e^{j\omega})}{\|\boldsymbol{A}(e^{j\omega})\|_{\Phi}^{2}} - \rho_{\Phi}(e^{j\omega})\frac{\boldsymbol{B}(e^{j\omega})}{\|\boldsymbol{A}(e^{j\omega})\|_{\Phi}\|\boldsymbol{B}(e^{j\omega})\|_{\Phi}}}{1-\left|\rho_{\Phi}(e^{j\omega})\right|^{2}} \quad (2.10)$$

where

$$\left\|\boldsymbol{X}(e^{j\omega})\right\|_{\Phi}^{2} \triangleq \boldsymbol{X}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}^{-1}(t,e^{j\omega})\boldsymbol{X}(e^{j\omega}) \quad (2.11)$$

denotes a weighted norm of a vector $\boldsymbol{X}(e^{j\omega})$ and

$$\rho_{\Phi}(e^{j\omega}) \triangleq \frac{\boldsymbol{B}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}^{-1}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})}{\sqrt{\boldsymbol{A}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}^{-1}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})}\sqrt{\boldsymbol{B}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}^{-1}(t,e^{j\omega})\boldsymbol{B}(e^{j\omega})}} \quad (2.12)$$

is the cosine of the angle between the vectors $\boldsymbol{A}(e^{j\omega})$ and $\boldsymbol{B}(e^{j\omega})$ in a weighted inner product space.

The closed-form solution to the constrained minimization problem, $W^{opt}(t,e^{j\omega})$, lacks the ability to track changes in the environment and is difficult to implement. Hence we replace the closed-form solution with an adaptive one. Let us consider the following steepest descent recursive algorithm for minimizing the complex Lagrangian in (2.8):

$$\mathbf{W}(t+1,e^{j\omega}) = \mathbf{W}(t,e^{j\omega}) - \mu\nabla_{W^{*}}\mathcal{L}(e^{j\omega}) =$$
$$\mathbf{W}(t,e^{j\omega}) - \mu\left[\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) + \lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega})\right]. \quad (2.13)$$

[3] As a small amount of uncorrelated sensor noise always exists, the invertibility of $\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})$ might be guaranteed in practical scenarios.

Imposing the constraints on $W(t+1, e^{j\omega})$ yields (see Appendix B):

$$\mathbf{W}(t+1, e^{j\omega}) = P(e^{j\omega})\mathbf{W}(t, e^{j\omega}) - \mu P(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t, e^{j\omega})\mathbf{W}(t, e^{j\omega}) + \mathbf{F}(e^{j\omega}) \quad (2.14)$$

where

$$P(e^{j\omega}) = I - \alpha^{-1}\left[\left\|\mathbf{B}(e^{j\omega})\right\|^2 \mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega}) - \mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega}) - \mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega}) + \left\|\mathbf{A}(e^{j\omega})\right\|^2 \mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\right]$$

$$\mathbf{F}(e^{j\omega}) = \alpha^{-1}\left[\left\|\mathbf{B}(e^{j\omega})\right\|^2 \mathbf{A}(e^{j\omega}) - \mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\mathbf{A}(e^{j\omega})\right]\mathcal{F}(e^{j\omega})$$

$$\alpha \triangleq \left\|\mathbf{A}(e^{j\omega})\right\|^2 \left\|\mathbf{B}(e^{j\omega})\right\|^2 - \mathbf{A}^{\dagger}(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\mathbf{A}(e^{j\omega}). \quad (2.15)$$

This forms the constrained recursive structure. Now, defining $\rho(e^{j\omega})$ as the coherence function (or the cosine of the angle between the vectors $\boldsymbol{A}(e^{j\omega})$ and $\boldsymbol{B}(e^{j\omega})$ in an inner product space)

$$\rho(e^{j\omega}) \equiv \frac{\boldsymbol{B}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\| \, \|\mathbf{B}(e^{j\omega})\|} \quad (2.16)$$

we obtain

$$P(e^{j\omega}) = I - \frac{\frac{\mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho^*(e^{j\omega})\frac{\mathbf{A}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|\|\mathbf{B}(e^{j\omega})\|} - \rho(e^{j\omega})\frac{\mathbf{B}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|\|\mathbf{B}(e^{j\omega})\|} + \frac{\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})}{\|\mathbf{B}(e^{j\omega})\|^2}}{1-|\rho(e^{j\omega})|^2}$$

$$\mathbf{F}(e^{j\omega}) = \frac{\frac{\mathbf{A}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho(e^{j\omega})\frac{\mathbf{B}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|\|\mathbf{B}(e^{j\omega})\|}}{1-|\rho(e^{j\omega})|^2}\mathcal{F}(e^{j\omega}). \quad (2.17)$$

The meaning of the coherence function will be discussed in the next section.

## 2.4 The dual source TFGSC

Following Gannot *et al.* [5] footsteps, we now derive an unconstrained adaptive enhancement algorithm. The unconstrained algorithm is usually advantageous due to its superior computational efficiency. Moreover, since the NLMS behavior is well explored in the literature, it is a preferable choice in many applications.

### 2.4.1 Generalized Sidelobe Canceller Interpretation

Consider the null space of $\left[\, \boldsymbol{A}(e^{j\omega}) \mid \boldsymbol{B}(e^{j\omega}) \,\right]$, defined by

$$\mathcal{N}(e^{j\omega}) \stackrel{\triangle}{=} \left\{ \boldsymbol{W} \mid \left[\, \boldsymbol{A}(e^{j\omega}) \mid \boldsymbol{B}(e^{j\omega}) \,\right]^{\dagger} \boldsymbol{W}(e^{j\omega}) = \begin{bmatrix} 0 \\ 0 \end{bmatrix} \right\}.$$

Define the constraint hyperplane,

$$\Lambda(e^{j\omega}) \stackrel{\triangle}{=} \left\{ \boldsymbol{W} \mid \left[\, \boldsymbol{A}(e^{j\omega}) \mid \boldsymbol{B}(e^{j\omega}) \,\right]^{\dagger} \boldsymbol{W}(e^{j\omega}) = \begin{bmatrix} \mathcal{F}(e^{j\omega}) \\ 0 \end{bmatrix} \right\}.$$

Since $\mathcal{N}(e^{j\omega})$ and $\Lambda(e^{j\omega})$ have the same perpendicular, while intersecting the axis in different points, $\Lambda(e^{j\omega})$ is parallel to $\mathcal{N}(e^{j\omega})$. Furthermore, define the column space of $\left[\, \boldsymbol{A}(e^{j\omega}) \mid \boldsymbol{B}(e^{j\omega}) \,\right]$ by

$$\mathcal{R}(e^{j\omega}) \stackrel{\triangle}{=} \left\{ \kappa_1 \boldsymbol{A}(e^{j\omega}) + \kappa_2 \boldsymbol{B}(e^{j\omega}) \mid \text{for any real } \kappa_1 \text{ and } \kappa_2 \right\}.$$

Using the fundamental theorem of Linear Algebra [20], $\mathcal{R}(e^{j\omega}) \perp \mathcal{N}(e^{j\omega})$. The second line in (2.17) implies that $\boldsymbol{F}(e^{j\omega}) \in \mathcal{R}(e^{j\omega})$ (as $\kappa_1$ and $\kappa_2$ can be easily identified) and therefore $\boldsymbol{F}(e^{j\omega})$ is perpendicular to $\mathcal{N}(e^{j\omega})$. Furthermore,

$$\boldsymbol{A}^{\dagger}(e^{j\omega})\boldsymbol{F}(e^{j\omega}) = \boldsymbol{A}^{\dagger}(e^{j\omega}) \frac{\frac{\mathbf{A}(\mathbf{e}^{\mathbf{j}\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho(e^{j\omega}) \frac{\mathbf{B}(\mathbf{e}^{\mathbf{j}\omega})}{\|\mathbf{A}(e^{j\omega})\| \|\mathbf{B}(e^{j\omega})\|}}{1 - |\rho(e^{j\omega})|^2} \mathcal{F}(e^{j\omega}) = \mathcal{F}(e^{j\omega})$$

and

$$\boldsymbol{B}^{\dagger}(e^{j\omega})\boldsymbol{F}(e^{j\omega}) = \boldsymbol{B}^{\dagger}(e^{j\omega}) \frac{\frac{\mathbf{A}(\mathbf{e}^{\mathbf{j}\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho(e^{j\omega}) \frac{\mathbf{B}(\mathbf{e}^{\mathbf{j}\omega})}{\|\mathbf{A}(e^{j\omega})\| \|\mathbf{B}(e^{j\omega})\|}}{1 - |\rho(e^{j\omega})|^2} \mathcal{F}(e^{j\omega}) = 0.$$

Hence, $\left[\, \boldsymbol{A}(e^{j\omega}) \mid \boldsymbol{B}(e^{j\omega}) \,\right]^{\dagger} \boldsymbol{F}(e^{j\omega}) = \begin{bmatrix} \mathcal{F}(e^{j\omega}) \\ 0 \end{bmatrix}$, i.e., $\boldsymbol{F}(e^{j\omega}) \in \Lambda(e^{j\omega})$. Now, since $\boldsymbol{F}(e^{j\omega}) \perp \mathcal{N}(e^{j\omega})$ and $\mathcal{N}(e^{j\omega})$ is parallel to $\Lambda(e^{j\omega})$, $\boldsymbol{F}(e^{j\omega}) \perp \Lambda(e^{j\omega})$. This implies that $\boldsymbol{F}(e^{j\omega})$ is the perpendicular from the origin to the constraint hyperplane, $\Lambda(e^{j\omega})$.

The matrix $P(e^{j\omega})$ is the *projection matrix* to the null space of $\left[\, \boldsymbol{A}(e^{j\omega}) \mid \boldsymbol{B}(e^{j\omega}) \,\right]$, $\mathcal{N}(e^{j\omega})$. This is easily shown by the following arguments. Using (2.17) we have,

$$P(e^{j\omega})\boldsymbol{A}(e^{j\omega}) = \boldsymbol{A}(e^{j\omega}) - \frac{1}{1 - |\rho(e^{j\omega})|^2} \left[ \frac{\boldsymbol{A}(e^{j\omega})\boldsymbol{A}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})}{\|\boldsymbol{A}(e^{j\omega})\|^2} - \right. \tag{2.18}$$

$$\left. \rho^*(e^{j\omega}) \frac{\boldsymbol{A}(e^{j\omega})\boldsymbol{B}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})}{\|\boldsymbol{A}(e^{j\omega})\| \, \|\boldsymbol{B}(e^{j\omega})\|} - \rho(e^{j\omega}) \frac{\boldsymbol{B}(e^{j\omega})\boldsymbol{A}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})}{\|\boldsymbol{A}(e^{j\omega})\| \, \|\boldsymbol{B}(e^{j\omega})\|} + \frac{\boldsymbol{B}(e^{j\omega})\boldsymbol{B}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})}{\|\boldsymbol{B}(e^{j\omega})\|^2} \right].$$

The term in the brackets is equal to

$$\begin{aligned}
&\boldsymbol{A}(e^{j\omega}) - \frac{1}{\|\boldsymbol{A}(e^{j\omega})\|^2 \|\boldsymbol{B}(e^{j\omega})\|^2} \left[\boldsymbol{A}^\dagger(e^{j\omega})\boldsymbol{B}(e^{j\omega})\boldsymbol{A}(e^{j\omega})\boldsymbol{B}^\dagger(e^{j\omega})\boldsymbol{A}(e^{j\omega}) - \right. \\
&\left. \boldsymbol{B}^\dagger(e^{j\omega})\boldsymbol{A}(e^{j\omega})\boldsymbol{B}(e^{j\omega})\boldsymbol{A}^\dagger(e^{j\omega})\boldsymbol{A}(e^{j\omega}) + \boldsymbol{B}^\dagger(e^{j\omega})\boldsymbol{A}(e^{j\omega})\boldsymbol{B}(e^{j\omega})\boldsymbol{A}^\dagger(e^{j\omega})\boldsymbol{A}(e^{j\omega})\right] \\
&= \boldsymbol{A}(e^{j\omega}) - \frac{\boldsymbol{A}^\dagger(e^{j\omega})\boldsymbol{B}(e^{j\omega})\boldsymbol{B}^\dagger(e^{j\omega})\boldsymbol{A}(e^{j\omega})}{\|\boldsymbol{A}(e^{j\omega})\|^2 \|\boldsymbol{B}(e^{j\omega})\|^2}\boldsymbol{A}(e^{j\omega}) = \boldsymbol{A}(e^{j\omega}) \left[1 - \left|\rho(e^{j\omega})\right|^2\right]
\end{aligned} \tag{2.19}$$

and therefore $P(e^{j\omega})\boldsymbol{A}(e^{j\omega}) = 0$. In a similar manner, $P(e^{j\omega})\boldsymbol{B}(e^{j\omega}) = 0$ as well.

Now, a vector in a linear space can be uniquely split into a sum of two vectors in mutually orthogonal subspaces (see for instance [20]). Hence,

$$\boldsymbol{W}(t, e^{j\omega}) = \boldsymbol{W}_0(t, e^{j\omega}) - \boldsymbol{V}(t, e^{j\omega}) \tag{2.20}$$

where $\boldsymbol{W}_0(t, e^{j\omega}) \in \mathcal{R}(e^{j\omega})$ and $-\boldsymbol{V}(t, e^{j\omega}) \in \mathcal{N}(e^{j\omega})$. By the definition of $\mathcal{N}(e^{j\omega})$,

$$\boldsymbol{V}(t, e^{j\omega}) = \mathcal{H}(e^{j\omega})\boldsymbol{G}(t, e^{j\omega}) \tag{2.21}$$

where $\mathcal{H}(e^{j\omega})$ is a full-rank $M \times (M-2)$ matrix, such that the columns of $\mathcal{H}(e^{j\omega})$ span the null space of $\left[\, \boldsymbol{A}(e^{j\omega}) \mid \boldsymbol{B}(e^{j\omega}) \,\right]$, i.e.,

$$\begin{aligned}
\boldsymbol{A}^\dagger(e^{j\omega})\mathcal{H}(e^{j\omega}) &= 0 \\
\boldsymbol{B}^\dagger(e^{j\omega})\mathcal{H}(e^{j\omega}) &= 0.
\end{aligned} \tag{2.22}$$

Note that due to signal space $\mathcal{R}(e^{j\omega})$ is of rank 2, there are exactly $(M-2)$ vectors in the null space. The vector $\boldsymbol{G}(t, e^{j\omega})$ is an $(M-2) \times 1$ vector of adjustable filters.

Using the geometrical interpretation of Frost's algorithm [1] (see Fig. 2.1), $\boldsymbol{W}_0(t, e^{j\omega})$ is the perpendicular from the origin to the constraint hyperplane $\Lambda(e^{j\omega})$, namely

$$\boldsymbol{W}_0(t, e^{j\omega}) = \boldsymbol{F}(e^{j\omega}) = \frac{\frac{\mathbf{A}(\mathbf{e}^{\mathbf{j}\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \rho(e^{j\omega})\frac{\mathbf{B}(\mathbf{e}^{\mathbf{j}\omega})}{\|\mathbf{A}(e^{j\omega})\|\|\mathbf{B}(e^{j\omega})\|}}{1 - \left\|\rho(e^{j\omega})\right|^2}\mathcal{F}(e^{j\omega}). \tag{2.23}$$

Now, using (2.4), (2.20) and (2.21) we obtain

$$Y(t, e^{j\omega}) = Y_{\mathrm{MBF}}(t, e^{j\omega}) - Y_{\mathrm{NC}}(t, e^{j\omega}) \tag{2.24}$$

where

$$Y_{\mathrm{MBF}}(t,e^{j\omega}) = \boldsymbol{W}_0^{\dagger}(t,e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}) \tag{2.25}$$
$$Y_{\mathrm{NC}}(t,e^{j\omega}) = \boldsymbol{G}^{\dagger}(t,e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}).$$

The solution structure is similar to [5], although the constraints are different. The output of the constrained beamformer is a difference of two terms, both operating on the input signal $\boldsymbol{Z}(t,e^{j\omega})$. $\boldsymbol{W}_0(t,e^{j\omega})$ in our problem, steers the beam towards the desired direction, while blocking the interference direction. In [5], $\boldsymbol{W}_0(t,e^{j\omega})$ is only responsible for steering the beam towards the desired direction. Furthermore, $\mathcal{H}(e^{j\omega})$ in the current contribution blocks both directions while in [5] it only blocks the desired direction. $\boldsymbol{G}(e^{j\omega})$ in both cases has similar functionality. However, its rank here is lower, allowing less degrees of freedom.

The first term, $Y_{\mathrm{MBF}}(t,e^{j\omega})$, is dependent on the ATFs, hence it can be regarded as a matched beamformer (MBF). We now examine the second term, $Y_{\mathrm{NC}}(t,e^{j\omega})$. The *reference noise* signals are given by

$$\begin{aligned}\boldsymbol{U}(t,e^{j\omega}) &= \mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}) \\ &= \mathcal{H}^{\dagger}(e^{j\omega})[\boldsymbol{A}(e^{j\omega})S_1(t,e^{j\omega}) + \boldsymbol{B}(e^{j\omega})S_2(t,e^{j\omega}) + \boldsymbol{N}(t,e^{j\omega})] = \mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{N}(t,e^{j\omega})\end{aligned} \tag{2.26}$$

where the last transition follows from (2.22). Both desired and competing signals' components are blocked by $\mathcal{H}^{\dagger}(e^{j\omega})$ and therefore $\boldsymbol{U}(t,e^{j\omega})$ contains only noise. Hence, the noise term of $Y_{\mathrm{MBF}}(t,e^{j\omega})$ can be reduced by properly adjusting the filters $\boldsymbol{G}(t,e^{j\omega})$, using the minimum output power criterion. This adjustment problem is in fact the classical multi-channel noise cancellation problem, that can be solved by using the Wiener filter. An adaptive LMS solution to the problem was proposed by Widrow [21].

Recall that the coherence function $\rho(e^{j\omega})$ defined in (2.16) is the cosine of the angle between $\boldsymbol{A}(e^{j\omega})$ and $\boldsymbol{B}(e^{j\omega})$. When these vectors are perpendicular $\rho(e^{j\omega})$ vanishes. In this case, the resulting $\mathbf{F}(e^{j\omega})$ is exactly the single source MBF derived in [5], and the projection matrix reduces to $P(e^{j\omega}) = I - \frac{\mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})}{\|\mathbf{A}(e^{j\omega})\|^2} - \frac{\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})}{\|\mathbf{B}(e^{j\omega})\|^2}$.

### 2.4.2 Detailed Structure

The solution comprises three building blocks. The first is an MBF, which satisfies the requested constraints, i.e. the desired signal is kept undistorted while the nonstationary interfering signal is blocked. The second is a blocking matrix, that produces noise-only reference signals by blocking both the desired and interfering signals. The third block is an unconstrained LMS-type algorithm, that cancels the coherent noise in the MBF output.

#### Blocking Matrix

The blocking matrix should be designed to block both the desired and interfering signals, and yield noise-only components at its outputs. We propose to construct $\mathcal{H}(e^{j\omega})$ as a cascade of two blocking matrices, $\mathcal{H}(e^{j\omega}) = \mathcal{H}_1(e^{j\omega})\mathcal{H}_2(e^{j\omega})$. $\mathcal{H}_1(e^{j\omega})$ is designed to block signals arriving from the desired signal direction, while $\mathcal{H}_2(e^{j\omega})$ is designed to block the signals arriving from the interfering direction, after being rotated by the first matrix. As in [5], $\mathcal{H}_1(e^{j\omega})$ is defined by

$$\mathcal{H}_1(e^{j\omega}) = \begin{bmatrix} -\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} & -\frac{A_3^*(e^{j\omega})}{A_1^*(e^{j\omega})} & \cdots & -\frac{A_M^*(e^{j\omega})}{A_1^*(e^{j\omega})} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \cdots & \ddots \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{2.27}$$

Regarding $\mathcal{H}_2(e^{j\omega})$, we have

$$\boldsymbol{B}^\dagger(e^{j\omega})\mathcal{H}(e^{j\omega}) = \boldsymbol{B}^\dagger(e^{j\omega})\left(\mathcal{H}_1(e^{j\omega})\mathcal{H}_2(e^{j\omega})\right) = \left(\boldsymbol{B}^\dagger(e^{j\omega})\mathcal{H}_1(e^{j\omega})\right)\mathcal{H}_2(e^{j\omega}). \tag{2.28}$$

Thus,

$$\boldsymbol{B}^\dagger(e^{j\omega})\mathcal{H}_1(e^{j\omega}) = \begin{bmatrix} B_1^*(e^{j\omega}) & B_2^*(e^{j\omega}) & \cdots & B_M^*(e^{j\omega}) \end{bmatrix} \cdot \begin{bmatrix} -\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} & -\frac{A_3^*(e^{j\omega})}{A_1^*(e^{j\omega})} & \cdots & -\frac{A_M^*(e^{j\omega})}{A_1^*(e^{j\omega})} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & & \cdots & \ddots \\ 0 & 0 & \cdots & 1 \end{bmatrix} = \begin{bmatrix} -B_1^*(e^{j\omega})\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_2^*(e^{j\omega}) \\ -B_1^*(e^{j\omega})\frac{A_3^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_3^*(e^{j\omega}) \\ \cdots \\ -B_1^*(e^{j\omega})\frac{A_M^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_M^*(e^{j\omega}) \end{bmatrix}^T.$$

This vector, multiplied by $\mathcal{H}_2(e^{j\omega})$, should yield a vector of zeros. Consider $\mathcal{H}_2(e^{j\omega})$ of the type

$$\mathcal{H}_2(e^{j\omega}) = \begin{bmatrix} L_3(e^{j\omega}) & L_4(e^{j\omega}) & \cdots & L_M(e^{j\omega}) \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix}. \tag{2.29}$$

The following linear equation determines $L_m(e^{j\omega});\ m = 3, \ldots, M$

$$\left[-B_1^*(e^{j\omega})\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_2^*(e^{j\omega})\right] L_m(e^{j\omega}) + \left[-B_1^*(e^{j\omega})\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} + B_m^*(e^{j\omega})\right] = 0. \tag{2.30}$$

Solving (2.30) we obtain:

$$L_m(e^{j\omega}) = -\frac{\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})}}{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}};\ m = 3, \ldots, M. \tag{2.31}$$

Multiplying $\mathcal{H}_1(e^{j\omega})$ by $\mathcal{H}_2(e^{j\omega})$ and rearranging terms yields

$$\mathcal{H}(e^{j\omega}) = \begin{bmatrix} Q_3(e^{j\omega}) & Q_4(e^{j\omega}) & \cdots & Q_M(e^{j\omega}) \\ L_3(e^{j\omega}) & L_4(e^{j\omega}) & \cdots & L_M(e^{j\omega}) \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{bmatrix} \tag{2.32}$$

where

$$Q_m(e^{j\omega}) = -\frac{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})}\frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})}}{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}};\ m = 3, \ldots, M. \tag{2.33}$$

We will now verify that $\mathcal{H}(e^{j\omega})$ satisfies both constraints in (2.22):

$$\begin{aligned} A^\dagger(e^{j\omega})\mathcal{H}(e^{j\omega}) = \big[&A_1^*(e^{j\omega})Q_3(e^{j\omega}) + A_2^*(e^{j\omega})L_3(e^{j\omega}) + A_3^*(e^{j\omega}); \\ &A_1^*(e^{j\omega})Q_4(e^{j\omega}) + A_2^*(e^{j\omega})L_4(e^{j\omega}) + A_4^*(e^{j\omega}); \cdots; \\ &A_1^*(e^{j\omega})Q_M(e^{j\omega}) + A_2^*(e^{j\omega})L_M(e^{j\omega}) + A_M^*(e^{j\omega})\big]. \end{aligned} \tag{2.34}$$

Calculating the $m$-th element in the last term

$$
\begin{aligned}
&A_1^*(e^{j\omega})Q_m(e^{j\omega}) + A_2^*(e^{j\omega})L_m(e^{j\omega}) + A_m^*(e^{j\omega}) \qquad (2.35)\\
&= \frac{1}{\frac{A_2^*(e^{j\omega})}{A_1^*(e^{j\omega})} - \frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})}} \cdot \Bigg\{ -A_2^*(e^{j\omega})\frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})} + A_m^*(e^{j\omega})\frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})} -\\
&A_2^*(e^{j\omega})\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} + A_2^*(e^{j\omega})\frac{B_m^*(e^{j\omega})}{B_1^*(e^{j\omega})} + A_2^*(e^{j\omega})\frac{A_m^*(e^{j\omega})}{A_1^*(e^{j\omega})} - A_m^*(e^{j\omega})\frac{B_2^*(e^{j\omega})}{B_1^*(e^{j\omega})} \Bigg\} = 0
\end{aligned}
$$

we obtain the required solution. Similarly, $\mathcal{H}(e^{j\omega})$ satisfies the second constraint, $B^{\dagger}(e^{j\omega})\mathcal{H}(e^{j\omega}) = 0$ and therefore is a valid blocking matrix which is suitable for generating the reference noise signals. By (2.26), we have

$$
\begin{aligned}
U_m(t, e^{j\omega}) &= Q_m(e^{j\omega})Z_1(t, e^{j\omega}) + L_m(e^{j\omega})Z_2(t, e^{j\omega}) + Z_m(t, e^{j\omega})\\
m &= 3, \ldots, M. \qquad (2.36)
\end{aligned}
$$

Thus, the knowledge of both $\frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{B_m(e^{j\omega})}{B_1(e^{j\omega})}$, or directly $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$, is sufficient for generating the noise reference signals. Note that the choice of $Z_1(t, e^{j\omega})$ and $Z_2(t, e^{j\omega})$ as the reference signals is arbitrary. In practical scenarios however, the references should be carefully selected amongst the input signals, and any more combinations thereof, in order to improve the estimation accuracy.

### Matched beamformer

It has been shown in Section 2.4.1 that the MBF $\boldsymbol{W}_0(e^{j\omega}) = \mathbf{F}(e^{j\omega})$, given by (2.17), satisfies the required constraints. If in (2.15), the actual ATFs are replaced by the ATFs ratios, then

$$
\boldsymbol{W}_0(e^{j\omega}) = \frac{\frac{\left\|\mathbf{B}(e^{j\omega})\right\|^2}{|B_1(e^{j\omega})|^2}\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})} - \frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\boldsymbol{B}^{\dagger}(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}}{\frac{\|\boldsymbol{A}(e^{j\omega})\|^2}{|A_1(e^{j\omega})|^2}\frac{\|\boldsymbol{B}(e^{j\omega})\|^2}{|B_1(e^{j\omega})|^2} - \frac{\boldsymbol{A}^{\dagger}(e^{j\omega})}{A_1^*(e^{j\omega})}\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\boldsymbol{B}^{\dagger}(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}}\mathcal{F}(e^{j\omega}). \qquad (2.37)
$$

When using the ATFs ratios instead of $\boldsymbol{A}(e^{j\omega})$ and $\boldsymbol{B}(e^{j\omega})$, we obtain:

$$\boldsymbol{W}_0^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega}) = \left[\frac{\frac{\left\|\mathbf{B}(e^{j\omega})\right\|^2}{|B_1(e^{j\omega})|^2}\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})} - \frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\boldsymbol{B}^{\dagger}(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}}{\frac{\|\boldsymbol{A}(e^{j\omega})\|^2}{|A_1(e^{j\omega})|^2}\frac{\|\boldsymbol{B}(e^{j\omega})\|^2}{|B_1(e^{j\omega})|^2} - \frac{\boldsymbol{A}^{\dagger}(e^{j\omega})}{A_1^*(e^{j\omega})}\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}\frac{\boldsymbol{B}^{\dagger}(e^{j\omega})}{B_1^*(e^{j\omega})}\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}}\right]^{\dagger}\mathcal{F}^*(e^{j\omega})\boldsymbol{A}(e^{j\omega})$$

$$= \frac{\frac{1}{|B_1(e^{j\omega})|^2}\frac{1}{A_1^*(e^{j\omega})}\left[\|\boldsymbol{B}(e^{j\omega})\|^2\|\boldsymbol{A}(e^{j\omega})\|^2 - \boldsymbol{A}^{\dagger}(e^{j\omega})\boldsymbol{B}(e^{j\omega})\boldsymbol{B}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})\right]^*}{\frac{1}{|A_1(e^{j\omega})|^2}\frac{1}{|B_1(e^{j\omega})|^2}\left[\|\boldsymbol{B}(e^{j\omega})\|^2\|\boldsymbol{A}(e^{j\omega})\|^2 - \boldsymbol{A}^{\dagger}(e^{j\omega})\boldsymbol{B}(e^{j\omega})\boldsymbol{B}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})\right]^*}\mathcal{F}^*(e^{j\omega})$$

$$= A_1(e^{j\omega})\mathcal{F}^*(e^{j\omega}). \tag{2.38}$$

Namely, the desired signal is only distorted by the first ATF $A_1(e^{j\omega})$, which can be absorbed into $\mathcal{F}^*(e^{j\omega})$. In a similar way, it can be shown that the nonstationary interference is completely blocked. Thus, the knowledge of the ATFs ratios is sufficient for implementing the sidelobe canceller.

It should be noticed that maximum directivity of the MBF can be obtained at directions other than the desired signal direction. Consider the next example, as depicted in Fig. 2.2. Polar plots of directivity patterns are computed for 5 and 10 microphones arrays for several frequencies in a simple delay-only ATFs system. The desired source signal is arriving from direction 90° while the non-stationary interference impinges on the array from direction 100°. It is clear that the beamformers satisfy both constraints in all plots, namely the gain is 1 in the desired direction and 0 in the interference direction. However, the array with 10 microphones outperforms the one with $M = 5$. For example, at 500Hz the array with 5 microphones has maximum gain of 3.3 at direction of $45^o$, while using 10 microphones maximum gain of 1.8 is achieved at direction of $75^o$, *i.e.* the peak is closer to the desired direction and to unity gain. Hence, interfering signals may be emphasized when not using enough microphones, especially in the low frequencies, which may deteriorate the ability of the ANC to cancel the stationary noise signals.

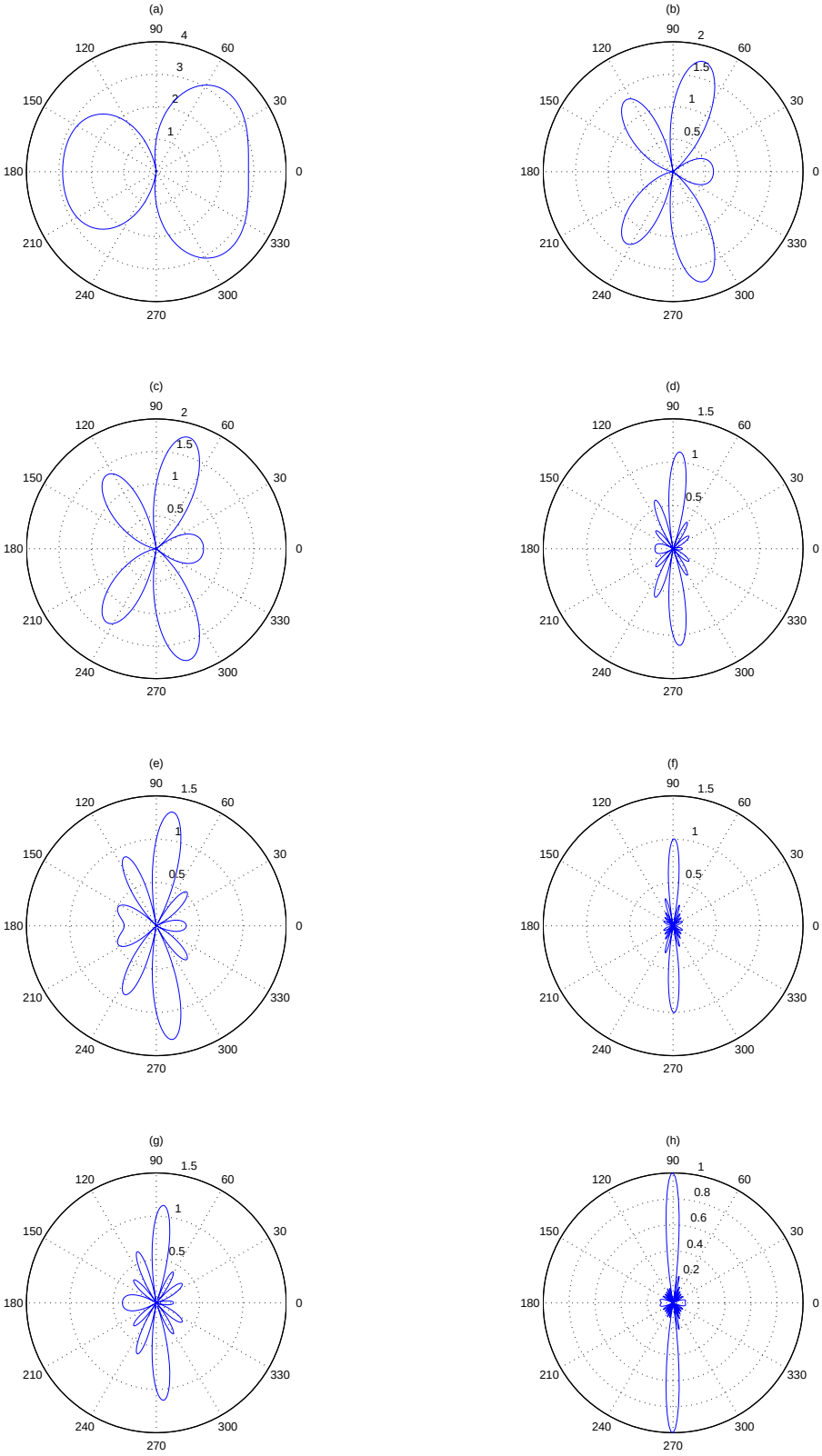


Figure 2.2: MBF directivity patterns for several scenarios: M=5, (a) f=500Hz (c) f=1000Hz (e) f=1500Hz (g) f=2000Hz; M=10, (b) f=500Hz (d) f=1000Hz (f) f=1500Hz (h) f=2000Hz.

**Multi channel noise canceller**

Recall that our goal is to minimize the output power under constraints on the response at the desired signal direction and at the competing signal direction. By setting $\boldsymbol{W}_0(t, e^{j\omega})$ according to (2.37), the constraints are satisfied. Hence, minimization of the output power is achieved by adjusting the filters $\boldsymbol{G}(t, e^{j\omega})$. This is an unconstrained minimization, exactly as in Widrow's classical problem [21]. It can be implemented by using the multi-channel Wiener filter. Recall (2.24), our goal is to set $\boldsymbol{G}(t, e^{j\omega})$ to minimize

$$\mathrm{E}\left\{||Y_{\mathrm{MBF}}(t, e^{j\omega}) - \boldsymbol{G}^{\dagger}(t, e^{j\omega})\boldsymbol{U}(t, e^{j\omega})||^2\right\}.$$

Let

$$\begin{aligned}\Phi_{\boldsymbol{U}Y}(t, e^{j\omega}) &= E\{\boldsymbol{U}(t, e^{j\omega})Y^*_{\mathrm{MBF}}(t, e^{j\omega})\}\\ \Phi_{\boldsymbol{U}\boldsymbol{U}}(t, e^{j\omega}) &= E\{\boldsymbol{U}(t, e^{j\omega})\boldsymbol{U}^{\dagger}(t, e^{j\omega})\}.\end{aligned}$$

Then the multi-channel Wiener filter is given by [22], [23]

$$\boldsymbol{G}(t, e^{j\omega}) = \Phi^{-1}_{\boldsymbol{U}\boldsymbol{U}}(t, e^{j\omega})\Phi_{\boldsymbol{U}Y}(t, e^{j\omega}). \tag{2.39}$$

In order to be able to track changes, the signals are processed by segments. The following frequency domain LMS algorithm is used. Let the residual signal be

$$Y(t, e^{j\omega}) = Y_{\mathrm{MBF}}(t, e^{j\omega}) - \boldsymbol{G}^{\dagger}(t, e^{j\omega})\boldsymbol{U}(t, e^{j\omega}).$$

Note that the residual signal is also the output of the enhancement algorithm. Using the orthogonality principle, the error is orthogonal to the measurements. Thus,

$$E\{\boldsymbol{U}(t, e^{j\omega})Y^*(t, e^{j\omega})\} = 0. \tag{2.40}$$

Following the standard Widrow procedure, the solution is given by:

$$\boldsymbol{G}(t+1, e^{j\omega}) = \boldsymbol{G}(t, e^{j\omega}) + \mu\boldsymbol{U}(t, e^{j\omega})Y^*(t, e^{j\omega}).$$

Usually, a more stable solution is obtained by using the normalized LMS (NLMS) algorithm, in which case each frequency is normalized separately, yielding:

$$G_m(t+1,e^{j\omega}) = G_m(t,e^{j\omega}) + \mu\frac{U_m(t,e^{j\omega})Y^*(t,e^{j\omega})}{P_{\text{est}}(t,e^{j\omega})}, \quad m=2,\ldots,M$$

where

$$P_{\text{est}}(t,e^{j\omega}) = \eta P_{\text{est}}(t-1,e^{j\omega}) + (1-\eta)\sum_m |Z_m(t,e^{j\omega})|^2 \tag{2.41}$$

and $\eta$ is a forgetting factor (typically $0.8 < \eta < 1$)[4].

The filter update is now given by

$$\begin{aligned} \tilde{G}_m(t+1,e^{j\omega}) &= G_m(t,e^{j\omega}) + \mu\frac{U_m(t,e^{j\omega})Y^*(t,e^{j\omega})}{P_{\text{est}}(t,e^{j\omega})} \\ G_m(t+1,e^{j\omega}) &\xleftarrow{\text{FIR}} \tilde{G}_m(t+1,e^{j\omega}) \end{aligned} \tag{2.42}$$

for $m=3,\ldots,M$. The operator $\xleftarrow{\text{FIR}}$ includes the following three stages. First, $\tilde{G}_m(t+1,e^{j\omega})$ is transformed to the time domain. Second, the resulting impulse response is truncated, namely an FIR constraint is imposed. Third, the result is transformed back to the frequency domain. Performing the $\xleftarrow{\text{FIR}}$ operator avoids cyclic convolution. A block diagram of the GSC solution is depicted in Fig. 2.3, and the proposed algorithm is summarized in Fig. 2.4.

## 2.5 ATFs Estimation

The ATFs ratios $\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}$ are required for calculating the MBF and the blocking matrix. Till this point, the ATFs were assumed to be known. However, in practice, they are estimated from the observed noisy signals. We assume that the ATFs ratios are slowly changing in time compared with the time variations of the desired signal and the competing speech signal. We also assume that the

[4]Another possibility is to calculate $P_{\text{est}}$ using the $|U_m(t,e^{j\omega})|^2$ instead of $|Z_m(t,e^{j\omega})|^2$. However, in that case an energy detector is required, so that $\boldsymbol{G}(t,e^{j\omega})$ is updated only when there is no active signal. If on the other hand, we calculate $P_{\text{est}}(t,e^{j\omega})$ using the input sensor signals, $Z_m(t,e^{j\omega})$, as indicated in (2.41), then an energy detector may be avoided. This is due to the fact that the adaptation term becomes relatively small during periods of active input signal.

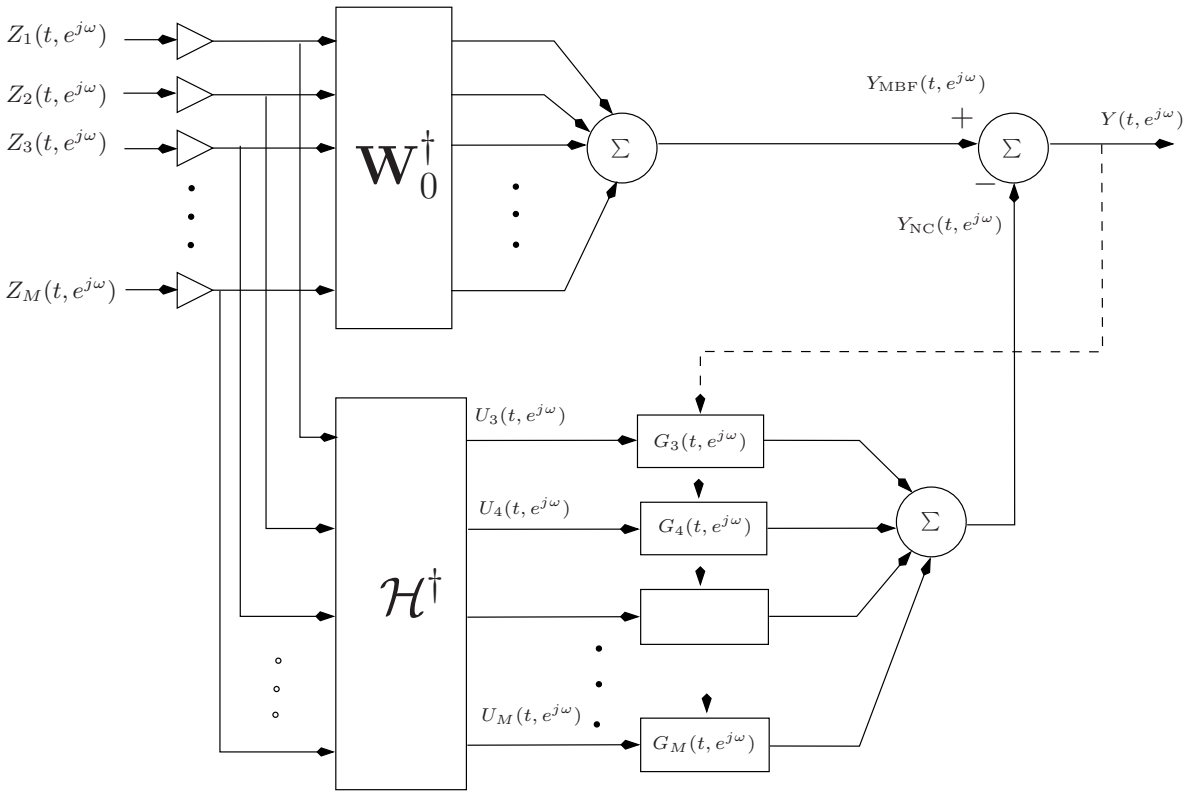


Figure 2.3: GSC solution for the dual source case. Three blocks: a matched beamformer $\boldsymbol{W}_0^\dagger(t,e^{j\omega})$; a blocking matrix $\mathcal{H}^\dagger(e^{j\omega})$; and a multi channel noise canceller $\boldsymbol{G}(t,e^{j\omega})$.

1) Matched beamformer:
$$Y_{\rm MBF}(t,e^{j\omega}) = \boldsymbol{W}_0^\dagger(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega})$$
2) Noise reference signals:
$$\boldsymbol{U}(t,e^{j\omega}) = \mathcal{H}^\dagger(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega})$$
3) Output signal:
$$Y(t,e^{j\omega}) = Y_{\rm MBF}(t,e^{j\omega}) - \boldsymbol{G}^\dagger(t,e^{j\omega})\boldsymbol{U}(t,e^{j\omega})$$
4) Filters update, for $m = 3,\ldots,M$:
$$\tilde{G}_m(t+1,e^{j\omega}) = G_m(t,e^{j\omega}) + \mu\frac{U_m(t,e^{j\omega})Y^*(t,e^{j\omega})}{P_{\rm est}(t,e^{j\omega})}$$
$$G_m(t+1,e^{j\omega}) \xleftarrow{\rm FIR} \tilde{G}_m(t+1,e^{j\omega})$$
where, $P_{\rm est}(t,e^{j\omega}) = \eta P_{\rm est}(t-1,e^{j\omega}) + (1-\eta)\sum_m |Z_m(t,e^{j\omega})|^2$
5) keep only non-aliased samples.
(note: $\boldsymbol{W}_0(e^{j\omega})$ is defined in (2.37). $\mathcal{H}(e^{j\omega})$ is defined in (2.32)).

Figure 2.4: Summary of the DTF-GSC algorithm.

statistics of the noise signal is slowly changing compared with the statistics of both the desired signal and the competing speech signal.

### 2.5.1 Matched Beamformer Estimate

Estimation of the MBF is carried out in two steps. First, the ATFs ratios $\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}$ are estimated separately, using the system identification procedure described in [5]. Second, $\boldsymbol{W}_0(t, e^{j\omega})$ is estimated using (2.37), where the ATFs ratios are used instead of the real ATFs. Since the system identification algorithm is designed for estimating a single system at a time, the two ratios cannot be estimated simultaneously. Therefore only frames in which both signals are not simultaneously active are used.

We will now briefly describe the system identification algorithm. The observation period is divided into frames such that the desired or the competing speech signals may be considered stationary during each $k$-th frame. Define $H_m(e^{j\omega}) \triangleq \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}$. Note that when no competing speech signal is in present, (2.2) turns to

$$Z_m(t, e^{j\omega}) = A_m(e^{j\omega})S(t, e^{j\omega}) + N_m(t, e^{j\omega}) \; ; \; m = 1, \ldots, M. \tag{2.43}$$

It is shown in [5] that the following $U_m(t, e^{j\omega})$ are proper noise reference signals,

$$U_m(t, e^{j\omega}) = Z_m(t, e^{j\omega}) - \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})} Z_1(t, e^{j\omega}) \; ; m = 2, \ldots, M. \tag{2.44}$$

Rearranging terms in (2.44) yields

$$Z_m(t, e^{j\omega}) = H_m(e^{j\omega})Z_1(t, e^{j\omega}) + U_m(t, e^{j\omega}). \tag{2.45}$$

Consider some analysis interval during which both the TF-s and the noise signal are assumed to be stationary. We divide that analysis interval into frames, such that the desired signal may be considered stationary during each frame. Consider the $k$-th frame, using (2.45) we have

$$\Phi^{(k)}_{z_m z_1}(e^{j\omega}) = H_m(e^{j\omega})\Phi^{(k)}_{z_1 z_1}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega}) \qquad k = 1, \ldots, K \tag{2.46}$$

where $K$ is the number of frames used and $\Phi^{(k)}_{z_i z_j}(e^{j\omega})$ is the cross-PSD between $z_i$ and $z_j$ during the $k$-th frame. $\Phi_{u_m z_1}(e^{j\omega})$ is the cross-PSD between $u_m$ and $z_1$. Using equations (2.43) and (2.44)

$$U_m(t, e^{j\omega}) = N_m(t, e^{j\omega}) - H_m(e^{j\omega}) N_1(t, e^{j\omega}) \tag{2.47}$$

$$Z_1(t, e^{j\omega}) = A_1(e^{j\omega}) S(t, e^{j\omega}) + N_1(t, e^{j\omega}) \tag{2.48}$$

Since $N_m(t, e^{j\omega}) = N_m(e^{j\omega})$ , $m = 1, \ldots, M$ are assumed stationary over the analysis interval and since $S(t, e^{j\omega})$ is independent of $N_m(e^{j\omega})$, it follows that $\Phi_{u_m z_1}(e^{j\omega})$ is independent of the frame index $k$.

Let $\hat{\Phi}^{(k)}_{z_1 z_1}(e^{j\omega})$, $\hat{\Phi}^{(k)}_{z_m z_1}(e^{j\omega})$ and $\hat{\Phi}^{(k)}_{u_m z_1}(e^{j\omega})$ be estimates of $\Phi^{(k)}_{z_1 z_1}(e^{j\omega})$, $\Phi^{(k)}_{z_m z_1}(e^{j\omega})$ and $\Phi_{u_m z_1}(e^{j\omega})$, respectively, and $\boldsymbol{U}(t, e^{j\omega})$ are the noise reference signals as defined in (2.44). The estimates are obtained by replacing expectations with averages. We then have,

$$\hat{\Phi}^{(k)}_{z_m z_1}(e^{j\omega}) = H_m(e^{j\omega}) \hat{\Phi}^{(k)}_{z_1 z_1}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega}) + \varepsilon^{(k)}_m(e^{j\omega})$$

where $\varepsilon^{(k)}_m(e^{j\omega}) = \hat{\Phi}^{(k)}_{u_m z_1}(e^{j\omega}) - \Phi_{u_m z_1}(e^{j\omega})$ denote the estimation error of the cross-PSD between $z_1$ and $u_m$ in the $k$-th frame. Hence, an unbiased estimate of $H_m(e^{j\omega})$ $(m = 2, \ldots, M)$ can be obtained by applying the least squares criterion to the following set of overdetermined equations

$$\begin{bmatrix} \hat{\Phi}^{(1)}_{z_m z_1}(e^{j\omega}) \\ \hat{\Phi}^{(2)}_{z_m z_1}(e^{j\omega}) \\ \vdots \\ \hat{\Phi}^{(K)}_{z_m z_1}(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} \hat{\Phi}^{(1)}_{z_1 z_1}(e^{j\omega}) & 1 \\ \hat{\Phi}^{(2)}_{z_1 z_1}(e^{j\omega}) & 1 \\ \vdots & \\ \hat{\Phi}^{(K)}_{z_1 z_1}(e^{j\omega}) & 1 \end{bmatrix} \begin{bmatrix} H_m(e^{j\omega}) \\ \Phi_{u_m z_1}(e^{j\omega}) \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)}_m(e^{j\omega}) \\ \varepsilon^{(2)}_m(e^{j\omega}) \\ \vdots \\ \varepsilon^{(K)}_m(e^{j\omega}) \end{bmatrix} \tag{2.49}$$

where a separate set of equations is used for $m = 2, \ldots, M$. Note that the ratio $\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}$ can be estimated in a similar manner.

### 2.5.2 Blocking Matrix Estimate

Inspecting (2.31) and (2.33), we note that the filters $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ can be estimated by using the ATFs ratios estimates and subsequently substituted into the blocking matrix given in (2.32). However, in this section we propose a direct

estimation method for $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$, which is applicable during double talk periods only. This facilitates tracking time variations of the blocking matrix when both the desired and interfering signals are active.

Rearranging terms in (2.36) yields

$$Z_m(t,e^{j\omega}) = -Q_m(e^{j\omega})Z_1(t,e^{j\omega}) - L_m(e^{j\omega})Z_2(t,e^{j\omega}) + U_m(t,e^{j\omega}). \tag{2.50}$$

We choose an observation period in which both the desired and competing speech signals are active simultaneously, and divide this period into frames such that the desired and directional interference signals may be considered stationary during each $k$-th frame. Using (2.50) we can obtain a system identification procedure,

$$\Phi^{(k)}_{z_m z_1}(e^{j\omega}) = -Q_m(e^{j\omega})\Phi^{(k)}_{z_1 z_1}(e^{j\omega}) - L_m(e^{j\omega})\Phi^{(k)}_{z_2 z_1}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega});\ k = 1,\ldots,K \tag{2.51}$$

where K is the number of frames in the interval, $\Phi^{(k)}_{z_i z_j}(e^{j\omega})$ is the *cross power spectral density* (cross-PSD) between $z_i$ and $z_j$ during the $k$-th frame, and $\Phi_{u_m z_1}(e^{j\omega})$ is the cross-PSD between $u_m$ and $z_1$ (note that $\Phi_{u_m z_1}(e^{j\omega})$ is independent of the frame index $k$ [5]). By replacing real PSD values with their estimates, calculated using time-averages, the following set of equations is obtained:

$$\begin{aligned}\hat{\Phi}^{(k)}_{z_m z_1}(e^{j\omega}) = &-Q_m(e^{j\omega})\hat{\Phi}^{(k)}_{z_1 z_1}(e^{j\omega}) - L_m(e^{j\omega})\hat{\Phi}^{(k)}_{z_2 z_1}(e^{j\omega}) + \hat{\Phi}_{u_m z_1}(e^{j\omega}) + \varepsilon^{(k)}_m(e^{j\omega}) \\ &k = 1,\ldots,K\end{aligned} \tag{2.52}$$

where $\varepsilon^{(k)}_m(e^{j\omega})$ represents the estimation error in the $k$-th frame. An unbiased estimate of $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ is obtained by applying the least squares criterion to the following over-determined set of equations

$$\begin{bmatrix} \hat{\Phi}^{(1)}_{z_m z_1}(e^{j\omega}) \\ \hat{\Phi}^{(2)}_{z_m z_1}(e^{j\omega}) \\ \vdots \\ \hat{\Phi}^{(K)}_{z_m z_1}(e^{j\omega}) \end{bmatrix} = \begin{bmatrix} \hat{\Phi}^{(1)}_{z_1 z_1}(e^{j\omega}) & \hat{\Phi}^{(1)}_{z_2 z_1}(e^{j\omega}) & 1 \\ \hat{\Phi}^{(2)}_{z_1 z_1}(e^{j\omega}) & \hat{\Phi}^{(2)}_{z_2 z_1}(e^{j\omega}) & 1 \\ \vdots & & \\ \hat{\Phi}^{(K)}_{z_1 z_1}(e^{j\omega}) & \hat{\Phi}^{(K)}_{z_2 z_1}(e^{j\omega}) & 1 \end{bmatrix} \times \begin{bmatrix} -Q_m(e^{j\omega}) \\ -L_m(e^{j\omega}) \\ \Phi_{u_m z_1}(e^{j\omega}) \end{bmatrix} + \begin{bmatrix} \varepsilon^{(1)}_m(e^{j\omega}) \\ \varepsilon^{(2)}_m(e^{j\omega}) \\ \vdots \\ \varepsilon^{(K)}_m(e^{j\omega}) \end{bmatrix} \tag{2.53}$$

where a separate set of equations is used for $m = 3,\ldots,M$.

It is worth noting that the proposed scheme enables estimation of $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ directly, and therefore facilitates blocking matrix update during double talk periods. We assume that perfect voice activity and double-talk detectors are available, and that in periods where only the desired signal is active, the estimate of $\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}$ is updated. This practical issue is beyond the scope of this work. However, in several applications, e.g. the joint echo cancellation and noise reduction problem, the detection of echo signal activity is trivial. We will elaborate on this issue in Chapter 4.

The latter estimate can be used in conjunction with the current estimate of $\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}$ to obtain the blocking matrix estimate during activity of the desired speech alone. The update of the blocking matrix by direct estimation of $Q_m(e^{j\omega})$ and $L_m(e^{j\omega})$ is done only during double talk periods, when no other update is possible. When the desired speech is inactive, the interfering and noise signals can be completely eliminated from the output of the algorithm. However, the BM and the MBF can be updated with the new estimates of $\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}$, to be used at the next active period of the desired signal.

## 2.6 Experimental results

The proposed algorithm was tested in a simulated room environment. The desired and competing speech signals were drawn from the TIMIT [24] database, while a speech-like noise from NOISEX-92 [25] database was used as a stationary noise source. All three signals were filtered by simulated room impulse responses, resulting in directional signals, which are received by M = 10 microphones. Allen and Berkley's *image method* [26] was used to simulate the ATFs. The parameter that is most often used to characterize reverberation is the reverberation time (also denoted $T_{60}$). This is defined as the time taken by a sound to decay 60 dB below the initial level, after it has been switched off [5]. The ATFs in this section

[5]The reverberation time can be calculated using the following formula: $T_{60} = \frac{-13.82}{c(L_x^{-1}+L_y^{-1}+L_z^{-1})ln(R)}$ where c is the speed of sound, $R$ is the reflection coefficient of all walls in the room and $L_x$, $L_y$, $L_z$ are the $x, y, z$ room dimensions, respectively.

were simulated with reverberation time set to $T_{60} = 200$ms (see Fig. 2.5 for a typical impulse and frequency responses of the acoustical path). In another test, diffused stationary noise was simulated [27] and used as the stationary noise. For the directional noise field scenario, the lengths of the filters in the MBF, the blocking matrix, and the interference cancellers are set to 250, 250 and 500 taps, respectively. For the diffused noise field scenario, the lengths of the filters are, respectively, 250, 250 and 300 taps. Segments of 1024 samples were used to implement the *overlap and save* procedure. The sampling frequency was set to 8KHz. The desired speech signal and the competing signal had the same level, with an average SNR of 5 dB.

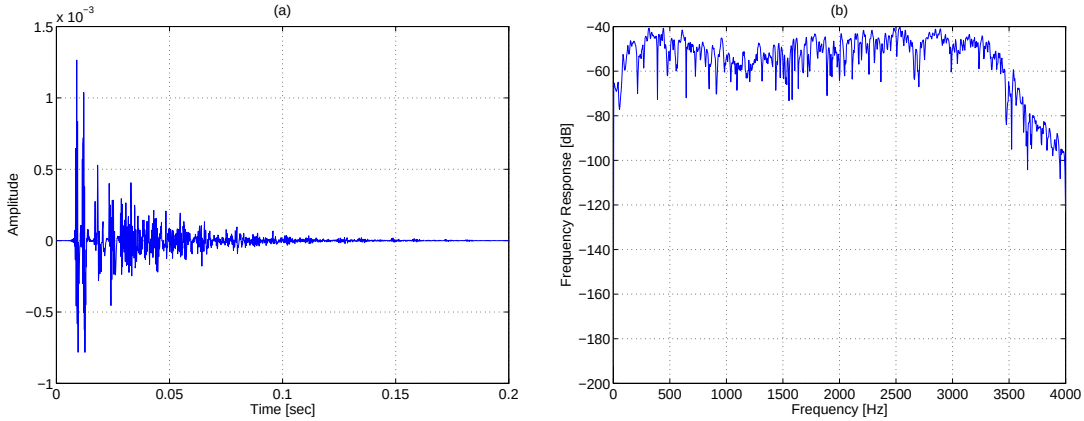


Figure 2.5: (a) Impulse response and (b) frequency response of the ATF between the desired source and the first microphone.

Fig. 2.6 shows the waveforms for the directional and diffused noise fields. In (a) and (b) a segment of the desired and competing speech signals are illustrated, respectively. Double talk situation is clearly observed. The signal measured at microphone # 1, which comprises the desired speech, the competing speech and a directional noise signal, is depicted in (c). The respective enhanced signal, after the algorithm has adapted, is depicted in (d). The noisy measurement at microphone # 1 and the enhanced signal for the diffused noise case are depicted in (e) and (f), respectively.

The SNR improvement for different stages of the GSC structure is given in Table 2.1, for both the directional and the diffused noise fields. Figures of merit given in Table 2.1 were calculated for the time segment depicted in Figures 2.6 and 2.7. $S_1NR$ and $S_2NR$ represent the signal to noise ratios for the signals $s_1(t)$ and $s_2(t)$, respectively. $S_1S_2R$ represents the ratio between the power of $s_1(t)$ and the power of $s_2(t)$. The results for the input signal at the first microphone, the outputs of the MBF and BM blocks, and total algorithm output are given in the Table. It is clearly seen that the MBF significantly improves the desired signal to interference ratio for both noise fields, while improving the desired signal to noise ratio as well. As expected, the BM deteriorates $S_1NR$ and $S_2NR$ for both noise fields, thus reducing the amount of desired signal leakage and allowing for the ANC to use almost noise-only reference signals. The noise level at the output is significantly reduced for the directional noise field case. The desired signal to noise level during double talk situation is increased by 24.5 dB. The desired signal to competing signal ratio level is increased by 11.9 dB. The noise reduction performance for the diffused noise field case is clearly inferior compared to the directional noise field case. The desired signal to noise level during double talk situation is improved by only 7.7 dB. The desired speech to the competing speech power ratio is increased by 11.0 dB, which is slightly inferior compared to the directional noise field case. Since the competing signal blocking ability depends on the ATFs estimation performance only, which is unrelated to the noise field type, no considerable difference is obtained.

Fig. 2.7 shows sonograms of the data depicted in Fig. 2.6. It can be seen that for the directional noise field, both noise and interference signals are well suppressed, especially in frequencies above 500 Hz. Moreover, no self cancellation or other deviation can be noticed during the double talk situation. Low noise reduction performance is obtained in the case of diffused noise field, while the nonstationary interference is well suppressed. The reduced noise reduction performance in the high frequencies (above 2500Hz) is due to the low coherence between the noise components in the received signals. Informal listening tests

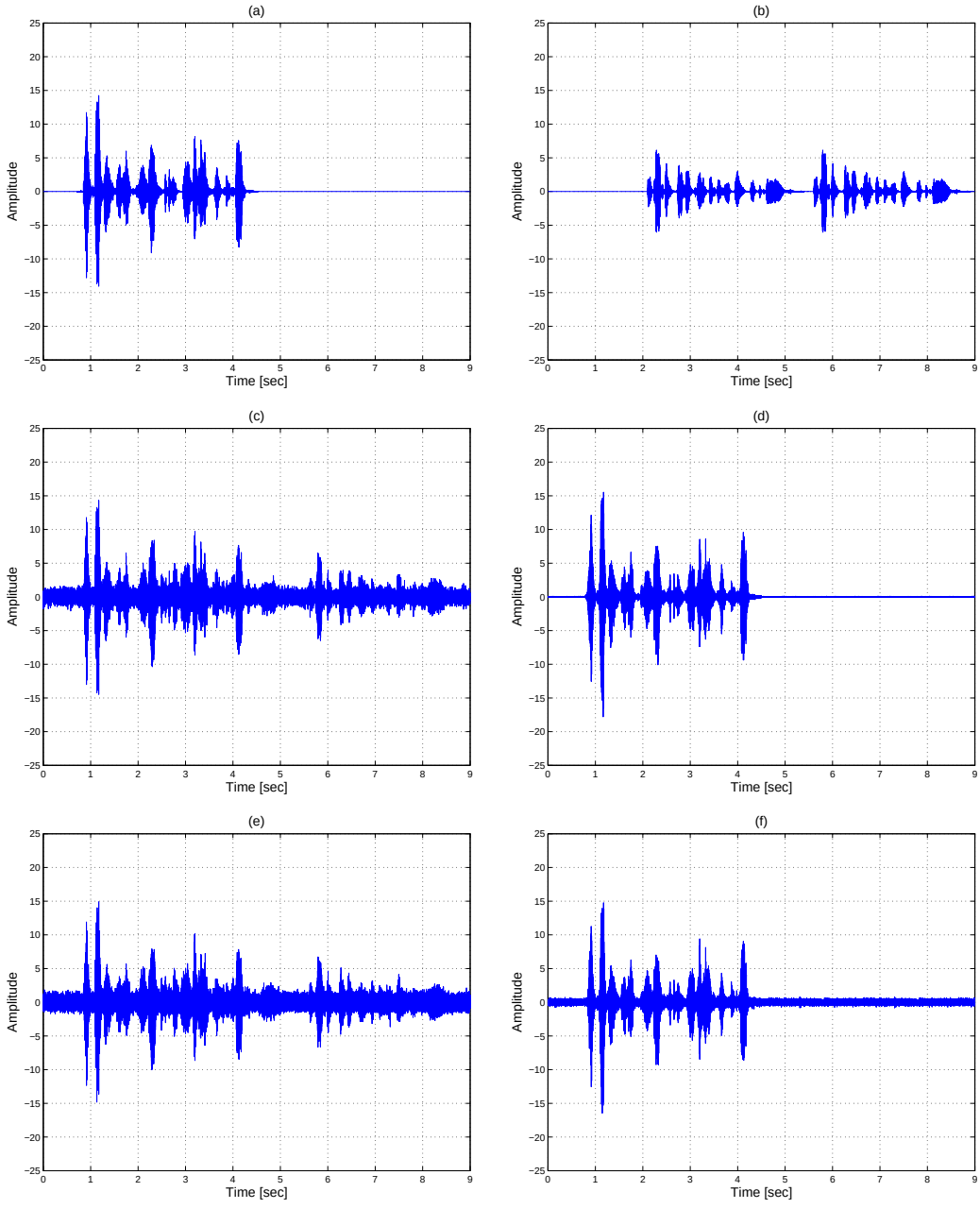


Figure 2.6: Speech waveforms: (a) Desired signal, (b) Nonstationary interference; Directional noise field: (c) Mic. #1 signal, (d) Enhanced signal; Diffused noise field: (e) Mic. #1 signal, (f) Enhanced signal.

Table 2.1: Noise and interference reduction in directional (top) and diffused (bottom) noise fields

| Noise Field | Input | | Output of MBF | | Output of BM | | Output of DTFGSC | |
|---|---|---|---|---|---|---|---|---|
| | $S_1NR$ | $S_1S_2R$ | $S_1NR$ | $S_1S_2R$ | $S_1NR$ | $S_2NR$ | $S_1NR$ | $S_1S_2R$ |
| Directional | 5.9 | 5.9 | 9.4 | -10.8 | 0.9 | -3.5 | 30.4 | 17.8 |
| Diffused | 7.1 | 5.9 | 11.0 | -9.2 | -0.5 | -4.9 | 14.8 | 16.9 |

confirm that the perceptual quality of the desired speech signal (for the directional noise field case) is retained in the enhanced signal, while the stationary and nonstationary interferences are well suppressed (audio sample files are available on-line [18]).

## 2.7 Summary

We presented a dual source interference canceller, based on the TF-GSC, for removing non-stationary directional interference and stationary interferences. The MBF and the blocking matrix were modified to handle the dual source case. A new system identification procedure was derived for estimating the blocking matrix terms directly during double talk segments. The proposed system may be applied to many interesting problems. One possible application is the BSS problem of convolutive mixtures and additive noise. The two sources can be extracted by exchanging the roles of the desired and competing speech signals. Another application is joint echo cancellation and noise reduction problem, obtained by replacing the competing speech with an echo signal. Note, however, that in this case the input echo signal is available and should be used to improve the obtained performance. This can be done, by incorporating the input echo signal as another input to the system, in a way similar to [15].

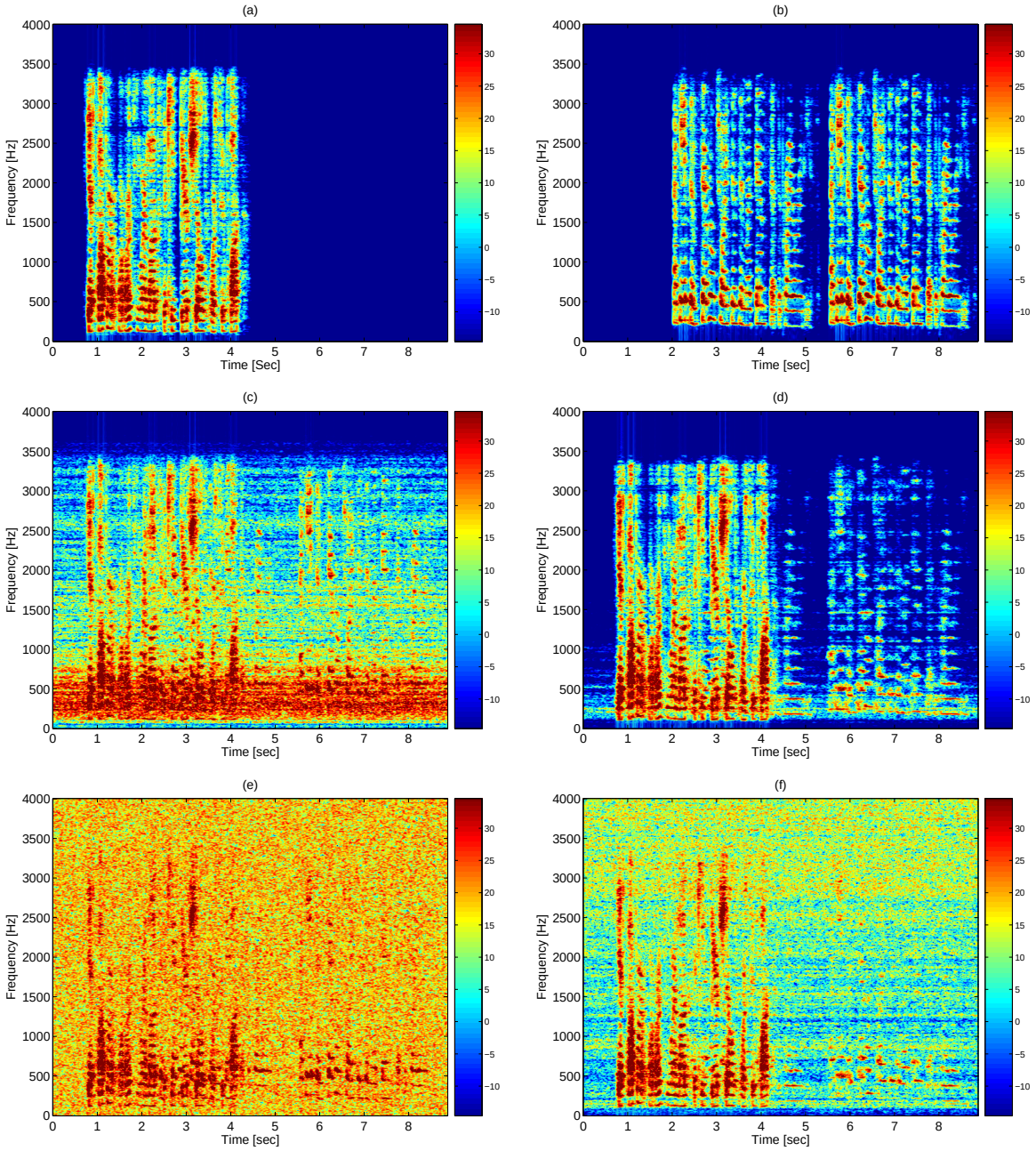


Figure 2.7: Sonograms: (a) Desired microphone #1, (b) Nonstationary interference microphone #1; directional noise field: (c) Noisy microphone #1, (d) Enhanced signal; diffused noise field: (e) Noisy microphone #1 and (f) Enhanced signal.

# Chapter 3

# Performance Analysis of the Dual Source TF-GSC

In this chapter we present a performance analysis of the DTF-GSC algorithm. The performance are evaluated in the presence of TFs' estimation errors, regardless of the estimation method. A general expression of the enhancer output is first derived. The expression depends on the TFs involved, the noise field and the quality of estimation of the TFs ratios. This analytical expression is used to evaluate three figures of merit, i.e. the amount of deviation imposed on the desired signal, the noise reduction and the interference reduction abilities.

## 3.1 Introduction

We consider a sensor array located in an enclosure, where arbitrary transfer functions (TFs) relate the source signal and the sensors. The array is used for enhancing a signal contaminated by interference signals. Gannot *et al.* [5] assumed the received signal is comprised of two components. The first is some nonstationary (e.g., speech) signal. The second is some stationary interference signal. The transfer function generalized sidelobe canceller (TF-GSC), a GSC solution which is adapted to the general TF case, was suggested to reconstruct the desired speech signal component from the received signals. A suboptimal algorithm that can be implemented by estimating the TFs ratios, was derived instead of estimating the

TFs themselves. The TF ratios are estimated by exploiting the nonstationarity characteristics of the desired signal. The discussion was supported by an experimental study using speech and noise signals recorded in an actual room acoustics environment.

Gannot *et al.* presented in [28] an analysis of a distortion indicator, namely power spectral density (PSD) deviation, imposed on the desired signal by the TF-GSC algorithm. It is well known that for speech signals, PSD deviation between the reconstructed signal and the original one is the main contribution for speech quality degradation. The resulting expression depends on the TFs involved, the noise field, and the quality of estimation of the TFs ratios. The theoretical expression is then used to establish empirical evaluation of the PSD deviation for several TFs of interest, various noise fields, and a wide range of SNR levels. It is shown that only a minor amount of PSD deviation is imposed on the beamformer output.

In Chapter 2 we considered a microphone array located in a reverberated room, where the desired signal is contaminated by directional nonstationary interference, such as a competing speaker, as well as stationary interferences which may comprise both directional and non-directional signals. The TF-GSC exploits the nonstationarity of the speech signal to enhance it when the undesired interfering signals are stationary. Unfortunately, the TF-GSC is rendered useless when a nonstationary interference is present. In Chapter 2, we propose an adaptive beamformer, based on the TF-GSC, that is suitable for cancelling nonstationary interferences in noisy reverberant environments. We modify two of the TF-GSC components to enable suppression of the nonstationary undesired signal. A modified fixed beamformer is designed to block the nonstationary interfering signal while maintaining the desired speech signal. A modified blocking matrix is designed to block both the desired signal and the nonstationary interference. We introduce a novel method for updating the blocking matrix in double talk situations, which exploits the nonstationarity of both the desired and interfering speech signals. Experimental results demonstrate the performance of the

proposed algorithm in noisy and reverberant environments.

Simulated speech environment, where the ATFs are modelled using *image method* [26], and speech signals are drawn from the TIMIT [24] database, demonstrates in Chapter 2 the ability of the algorithm to reduce noise and interference signals, while maintaining the desired signal content. Nevertheless, it is also interesting to perform analytical evaluation of the expected performance, especially for determining the performance limits. Gannot *et al.* [28] presented an analysis of the PSD deviation imposed on the desired signal by the TF-GSC algorithm. However, the structure of the matched beamformer (MBF) and blocking matrix (BM) in the DTF-GSC differs from the TF-GSC since they both constrained to block the competing speech signal direction. We show in Chapter 2 that the deviation considerably depends on the distance between the desired source and remote source, due to the contradicted constraints imposed on the MBF. Hence, the PSD deviation should be reevaluated for the DTF-GSC algorithm. In addition, the DTF-GSC performs interference reduction as well as noise reduction and both abilities should be also analyzed.

The goal of this chapter is to present a comprehensive expression for the output of the DTF-GSC structure when no restrictions are imposed on the TFs relating the source and the sensors. From the general expression, the noise reduction, interference reduction and PSD deviation can be derived. We show that the resulting expressions for the PSD deviation and noise reduction depend on the actual TFs, the noise field type, and the accuracy of the TFs ratio estimation. However, exact knowledge of the TFs' ratio results in complete interference reduction, regardless of the noise field type.

The structure of this chapter is as follows. We begin by briefly reviewing the DTF-GSC algorithm in Section 3.2. In Section 3.3 we formulate a general expression for the algorithm's output power spectral density. The desired signal PSD deviation is evaluated in Section 3.4. Evaluations of the noise reduction performance and its dependency on the noise field is addressed in Section 3.5. The achievable interference reduction is analyzed in Section 3.6.

## 3.2 Summary of the DTF-GSC

In Chapter 2 we considered a microphone array located in a reverberated room, where general transfer functions (TFs) relate the source signal and the microphones. The array is used for enhancing a speech signal contaminated by a competing speech signal as well as stationary noise signal, i.e. two nonstationary sources are present. The new algorithm enables reconstruction of the desired speech signal (allowing reverberations) from the contaminated observations.

A block diagram of the DTF-GSC solution is depicted in Fig. 3.1, and the algorithm is summarized in Fig. 3.2.

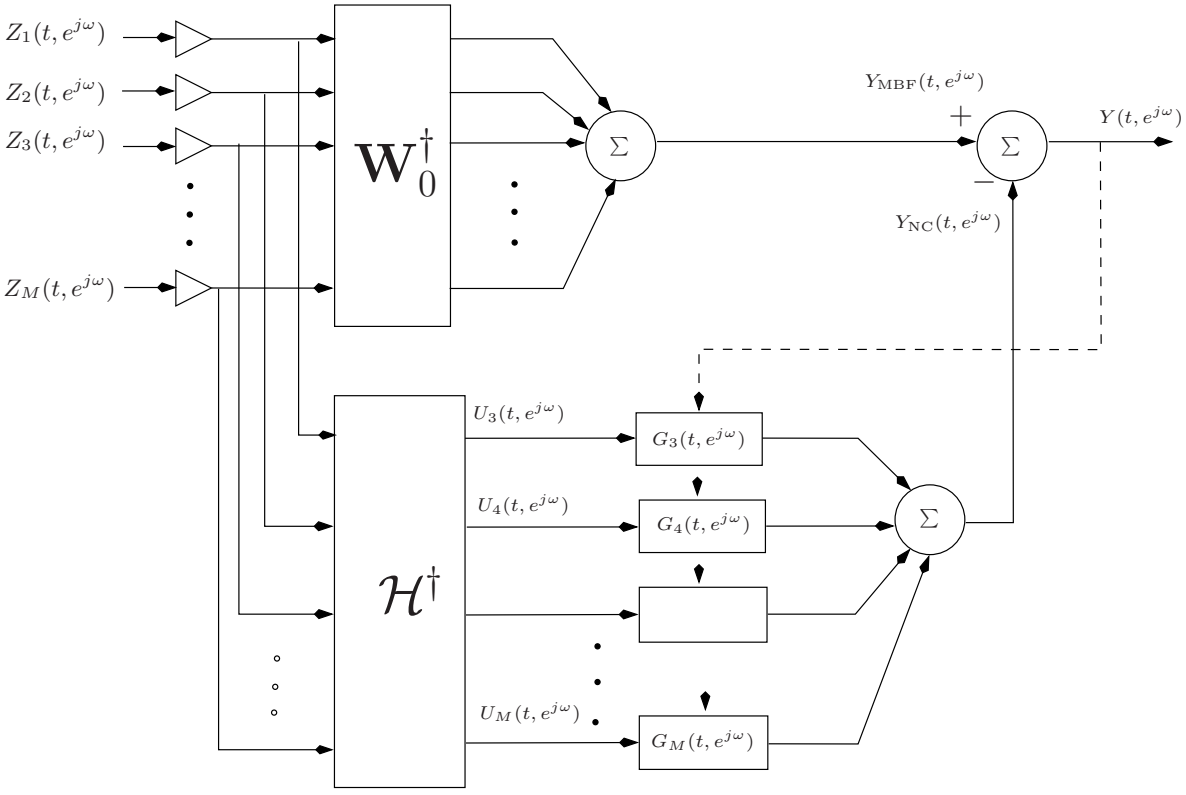


Figure 3.1: GSC solution for the dual source case. Three blocks: a matched beamformer $\boldsymbol{W}_0^\dagger(t,e^{j\omega})$; a blocking matrix $\mathcal{H}^\dagger(e^{j\omega})$; and a multi channel noise canceller $\boldsymbol{G}(t,e^{j\omega})$.

1) Matched beamformer:
$$Y_{\text{MBF}}(t, e^{j\omega}) = \boldsymbol{W}_0^{\dagger}(e^{j\omega})\boldsymbol{Z}(t, e^{j\omega})$$
2) Noise reference signals:
$$\boldsymbol{U}(t, e^{j\omega}) = \mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t, e^{j\omega})$$
3) Output signal:
$$Y(t, e^{j\omega}) = Y_{\text{MBF}}(t, e^{j\omega}) - \boldsymbol{G}^{\dagger}(t, e^{j\omega})\boldsymbol{U}(t, e^{j\omega})$$
4) Filters update, for $m = 3, \ldots, M$:
$$\tilde{G}_m(t+1, e^{j\omega}) = G_m(t, e^{j\omega}) + \mu \frac{U_m(t, e^{j\omega}) Y^*(t, e^{j\omega})}{P_{\text{est}}(t, e^{j\omega})}$$
$$G_m(t+1, e^{j\omega}) \xleftarrow{\text{FIR}} \tilde{G}_m(t+1, e^{j\omega})$$
where, $P_{\text{est}}(t, e^{j\omega}) = \eta P_{\text{est}}(t-1, e^{j\omega}) + (1-\eta)\sum_m |Z_m(t, e^{j\omega})|^2$
5) keep only non-aliased samples.
(note: $\boldsymbol{W}_0(e^{j\omega})$ is defined in (2.37). $\mathcal{H}(e^{j\omega})$ is defined in (2.32)).

Figure 3.2: Summary of the DTF-GSC algorithm.

Both the TF-GSC and the DTF-GSC share the same GSC structure. Nevertheless, two of the components are modified, namely the fixed beamformer and the blocking matrix, to allow the beamformer to suppress the nonstationary interference. The modified fixed beamformer is designed to block the nonstationary interference while maintaining the desired speech signal. The modified BM blocks both the desired signal and the nonstationary interference. As in the TF-GSC structure, the adaptive noise canceller employs the reference signals generated by the blocking matrix to cancel the residual stationary noise at the MBF output. It was shown that it is sufficient to use the ratio between the different TFs rather than the TFs themselves in order to implement the suggested algorithm. Novel approach was proposed to estimate the blocking matrix during double talk frames. Experimental results show that the algorithm is working well in various real-life scenarios.

We show in Chapter 2 that the deviation considerably depends on the distance between the desired source and remote source, due to the contradicting constraints imposed on the MBF. Furthermore, the BM in the DTF-GSC imposes two constraints while the BM in the TF-GSC imposes only one, using the same number of microphones. Hence, the PSD deviation and the noise reduc-

tion performance should be reevaluated for the DTF-GSC algorithm, while the interference reduction ability should be analyzed.

## 3.3 Output Power Spectral Density

A general expression of the enhancer output PSD is essential to evaluate the performance. Using (2.24) and (2.25) the algorithm's output is given by,

$$\begin{aligned} Y(t,e^{j\omega}) &= Y_{\mathrm{MBF}}(t,e^{j\omega}) - \boldsymbol{G}^{\dagger}(t,e^{j\omega})\boldsymbol{U}(t,e^{j\omega}) \\ &= \widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}) - \boldsymbol{G}^{\dagger}(t,e^{j\omega})\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}), \end{aligned} \tag{3.1}$$

where only estimates of the TFs ratio, rather than the exact values, are assumed to be known. Using this expression, the PSD of the output signal can be calculated,

$$\begin{aligned} \Phi_{yy}(t,e^{j\omega}) &= E\{Y(t,e^{j\omega})Y^{*}(t,e^{j\omega})\} \\ &= E\left\{\left(\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}) - \boldsymbol{G}^{\dagger}(t,e^{j\omega})\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega})\right)\right. \\ &\quad \left.\times\left(\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega}) - \boldsymbol{G}^{\dagger}(t,e^{j\omega})\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t,e^{j\omega})\right)^{\dagger}\right\}. \end{aligned} \tag{3.2}$$

Opening brackets and using the PSD definition $\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega}) = E\{\boldsymbol{Z}(t,e^{j\omega})\boldsymbol{Z}^{\dagger}(t,e^{j\omega})\}$ yields,

$$\begin{aligned} \Phi_{yy}(t,e^{j\omega}) = {} & \widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \\ & -\boldsymbol{G}^{\dagger}(t,e^{j\omega})\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \\ & -\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\boldsymbol{G}(t,e^{j\omega}) \\ & +\boldsymbol{G}^{\dagger}(t,e^{j\omega})\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\boldsymbol{G}(t,e^{j\omega}). \end{aligned} \tag{3.3}$$

The output PSD depends on the input signal $\boldsymbol{Z}(t,e^{j\omega})$ and the optimal multichannel Wiener filter given by (2.39). Although signal leakage might occur due to errors in estimating $\frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{B_m(e^{j\omega})}{B_1(e^{j\omega})}$, we can calculate the filter during non-active periods of the desired signal and nonstationary interference, in which

$\boldsymbol{Z}(t, e^{j\omega}) = \boldsymbol{N}(t, e^{j\omega})$. Thus,

$$\begin{aligned} \boldsymbol{U}(t, e^{j\omega})\big|_{Z=N} &= \widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\boldsymbol{N}(t, e^{j\omega}) \\ Y_{\mathrm{MBF}}|_{Z=N} &= \widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\boldsymbol{N}(t, e^{j\omega}). \end{aligned}$$

Calculating the involved PSDs

$$\begin{aligned} \Phi_{\boldsymbol{U}Y_{\mathrm{MBF}}}(t, e^{j\omega}) &= E\{\boldsymbol{U}(t, e^{j\omega})Y^{*}_{\mathrm{MBF}}(t, e^{j\omega})\} \\ &= E\left\{\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\boldsymbol{N}(t, e^{j\omega})\left(\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\boldsymbol{N}(t, e^{j\omega})\right)^{\dagger}\right\} \\ &= \widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \end{aligned} \tag{3.4}$$

and

$$\begin{aligned} \Phi_{\boldsymbol{UU}}(t, e^{j\omega}) &= E\{\boldsymbol{U}(t, e^{j\omega})\boldsymbol{U}^{\dagger}(t, e^{j\omega})\} \\ &= E\left\{\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\boldsymbol{N}(t, e^{j\omega})\boldsymbol{N}^{\dagger}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right\} \\ &= \widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}), \end{aligned} \tag{3.5}$$

where $\Phi_{\boldsymbol{NN}}(t, e^{j\omega}) = E\{\boldsymbol{N}(t, e^{j\omega})\boldsymbol{N}^{\dagger}(t, e^{j\omega})\}$, we obtain,

$$\boldsymbol{G}(t, e^{j\omega}) = \left(\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1}\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}). \tag{3.6}$$

Note that $\boldsymbol{G}(t, e^{j\omega})$ deals with the noise signal only, while the competing speech signal $s_2(t)$ is blocked by the MBF and BM.

Let us define $T_1$ as the time interval in which $s_1(t)$ is the only nonstationary signal active, while $T_2$ is the time interval in which $s_2(t)$ is the only nonstationary signal active. $T_n$ is defined as the time interval in which only noise signal is present. Hence, using desired signal, interference signal and noise signal independence, the *noise reduction*, *interference reduction* and the *deviation* imposed by the algorithm, can be calculated separately by deriving expressions for the output PSD in the following three situations:

$$\Phi_{\boldsymbol{ZZ}}(t, e^{j\omega}) = \begin{cases} \Phi_{\boldsymbol{NN}}(t, e^{j\omega}) & t \in T_n \\ \Phi_{s_1s_1}(t, e^{j\omega})\boldsymbol{A}(e^{j\omega})\boldsymbol{A}^{\dagger}(e^{j\omega}) & t \in T_1 \\ \Phi_{s_2s_2}(t, e^{j\omega})\boldsymbol{B}(e^{j\omega})\boldsymbol{B}^{\dagger}(e^{j\omega}) & t \in T_2 \end{cases}$$

yielding,

$$\Phi_{yy}(t,e^{j\omega}) = \begin{cases} \Phi^{n}_{yy}(t,e^{j\omega}) \ t \in T_n \Rightarrow & \text{noise reduction} \\ \Phi^{s_1}_{yy}(t,e^{j\omega}) \ t \in T_1 \Rightarrow & \text{desired signal deviation} \\ \Phi^{s_2}_{yy}(t,e^{j\omega}) \ t \in T_2 \Rightarrow & \text{interference reduction} \end{cases}$$

Using (3.6) and (3.3) we obtain the output signal output PSD given in (3.7).

$$\begin{aligned} \Phi_{yy}(t,e^{j\omega}) = & \Big\{ \widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \\ & -\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) \left(\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \\ & -\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) \left(\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \\ & +\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) \left(\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) \\ & \times \left(\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \ \Big\} . \end{aligned} \tag{3.7}$$

This complicated expression depends on various parameters: the input signal PSD, $\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})$, the noise PSD used for calculating the optimal filters, $\Phi_{\boldsymbol{NN}}(t,e^{j\omega})$ and the ATFs' ratio $\frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{B_m(e^{j\omega})}{B_1(e^{j\omega})}$. Gannot *et al.* [28] presented similar expression for evaluating the TF-GSC deviation performance, although $\widehat{\boldsymbol{W_0}}(e^{j\omega})$ and $\widehat{\mathcal{H}}(e^{j\omega})$ are completely different in the DTF-GSC.

Expression (3.7) will be used is Sections 3.4,3.5 and 3.6 for deriving general expressions for the deviation imposed by the DTF-GSC algorithm, noise reduction and interference reduction achieved, respectively.

## 3.4 Evaluation of Desired Signal PSD Deviation

The desired signal PSD imposed by the algorithm can be calculated by the general expression given in (3.7) for a signal $\boldsymbol{Z}(t,e^{j\omega}) = \boldsymbol{A}(e^{j\omega})s_1(t,e^{j\omega})$. Assume we have exact knowledge of the TFs' ratio $\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}$, i.e. $\widehat{\mathcal{H}}(e^{j\omega}) = \mathcal{H}(e^{j\omega})$ and $\widehat{\boldsymbol{W_0}}(e^{j\omega}) = \boldsymbol{W_0}(e^{j\omega})$, calculated using (2.31), (2.33), (2.32) and (2.37). Thus, using the signal PSD expression $\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega}) = \Phi_{s_1s_1}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})\boldsymbol{A}^{\dagger}(e^{j\omega})$ and

the identities $\mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega}) = 0$, $\mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{B}(e^{j\omega}) = 0$ and $\boldsymbol{A}^{\dagger}(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}) = A_1(e^{j\omega})\mathcal{F}(e^{j\omega})$, expression (3.7) reduces to

$$\begin{aligned}\Phi_{yy}^{s_1}(t,e^{j\omega})\big|_{\widehat{H}=H} &= \boldsymbol{W_0}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{W_0}(e^{j\omega}) = \Phi_{s_1s_1}(t,e^{j\omega})|\boldsymbol{W_0}^{\dagger}(e^{j\omega})\boldsymbol{A}(e^{j\omega})|^2 \\ &= \Phi_{s_1s_1}(t,e^{j\omega})|\mathcal{F}(e^{j\omega})|^2|A_1(e^{j\omega})|^2. \end{aligned} \tag{3.8}$$

The filter $\mathcal{F}(e^{j\omega})$ is an arbitrary predetermined filter, so it should not be regarded as a deviation. The filter $A_1(e^{j\omega})$ is the TF from the signal source to the first (reference) sensor. It cannot be eliminated by the algorithm. Actually, it imposes on the output signal the same amount of PSD imposed on the arbitrary reference sensor. Thus, we define the deviation from the nominal value of the algorithm by normalizing the output,

$$\mathrm{DEV}(t,e^{j\omega}) = \frac{\Phi_{yy}^{s_1}(t,e^{j\omega})}{|\mathcal{F}(e^{j\omega})|^2|A_1(e^{j\omega})|^2\Phi_{s_1s_1}(t,e^{j\omega})}. \tag{3.9}$$

A value of $\mathrm{DEV}(t,e^{j\omega}) = 0dB$ indicates no deviation from the nominal value is obtained. This value is achieved whenever an exact knowledge of the TFs ratios is available. Any other value, negative or positive, means that the received signal is attenuated or amplified, respectively.

We will determine now the influence of errors in estimating the TFs on the expected deviation imposed by the algorithm. The results depend on the desired signal's TFs and the noise field. We begin with the simple steering array, where only delays relate the sources and the sensors. Thus, the *direction of arrival* (DoA) of the sources completely determines the TFs. We determine the performance degradation as a function of the steering error. Then, we deal with the real room case.

### 3.4.1 Signal's TFs: Pure Delay

Assume free space propagation, i.e., the ATFs from the desired source and interference source to the sensors are pure delay. Assume that the desired signal impinges on the array from $\theta = 90^o$ and that the inter-element distance is 10cm.

We will examine two scenarios. In the first, denoted *remote sources*, the non-stationary interference impinges on the array from $\theta = 60^o$, while in the second, denoted *close sources*, $\theta = 85^o$. The deviation imposed by the algorithm is evaluated by steering the array to the assumed desired signal DoA ($\theta = 90^o$) while changing the real direction of the signal in the range $\theta \in [87^o, 93^o]$. Thus the interference signal is close to the desired signal.

**Directional Noise Signal**

In the directional noise field case we optimize the array to cancel a noise source from $\theta = 120^o$ (by optimization of the array we refer to designing the optimal Wiener filter in the noise cancellation branch). In Figure 3.3 (a) and (b) we present the deviation of the output signal of the array as a function of the frequency and the steering angle, for the remote sources scenario and close sources scenario, respectively. It is clearly shown that for both scenarios the output signal is distorted as we move the steering angle of the array away from the desired signal direction $\theta = 90^o$.

In the remote sources scenario, the imposed deviation is not exceeding -4dB for frequencies upto 3000Hz, where maximum value is 0dB and obtained in the desired source direction, i.e. $\theta = 90^o$. On the other hand, in the close scenario, the deviation might reach -8dB, where maximum value of 0dB at $\theta = 90^o$ is obtained only in high frequencies. In fact, for low frequencies the maximum deviation is obtained in a direction which is outside the plot axis.

The spatial beamformer response in 500Hz and 3000Hz is depicted in Figure 3.4 (a) and (b), for the remote and close scenarios, respectively. It can be clearly seen that the main lobe around the desired direction is wider at 500Hz than in 3000Hz. This is in good agreement with the theory, since at $\omega = 0$ [rad/sec] there is no phase difference between the signals at the sensors. Wide main lobe means that error in the estimated direction of arrival, causes only small deviation in the desired signal. In this case, only negligible impairment to the desired signal is obtained.

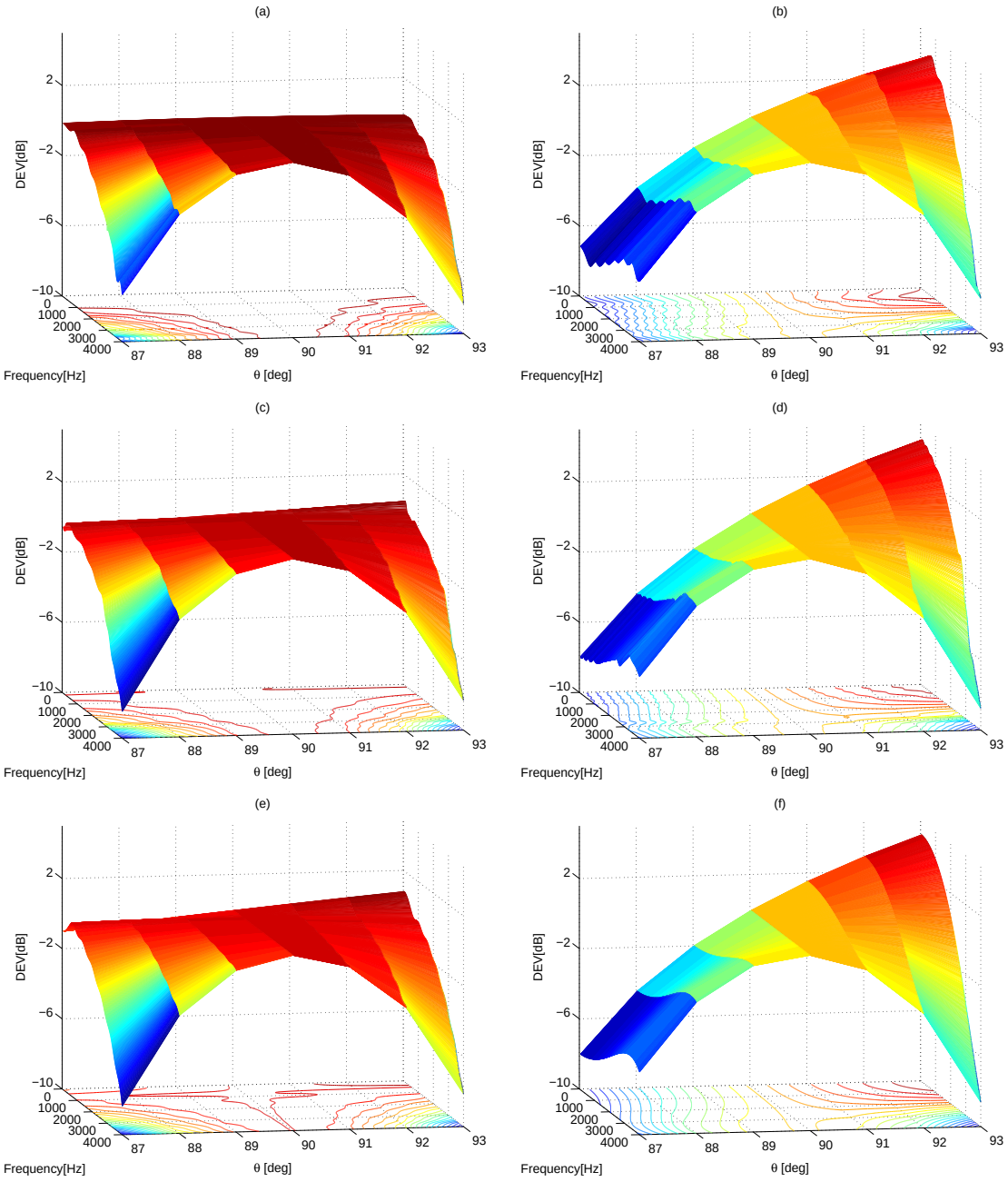


Figure 3.3: Deviation as a function of the frequency and direction of arrival. Desired signal direction $\theta = 90^o$. $M = 10$ sensors. remote sources: (a) Directional noise field ($\theta = 120^o$), (c) Diffused noise field, (e) Incoherent noise field; close sources: (b) Directional noise field ($\theta = 120^o$), (d) Diffused noise field, (f) Incoherent noise field.

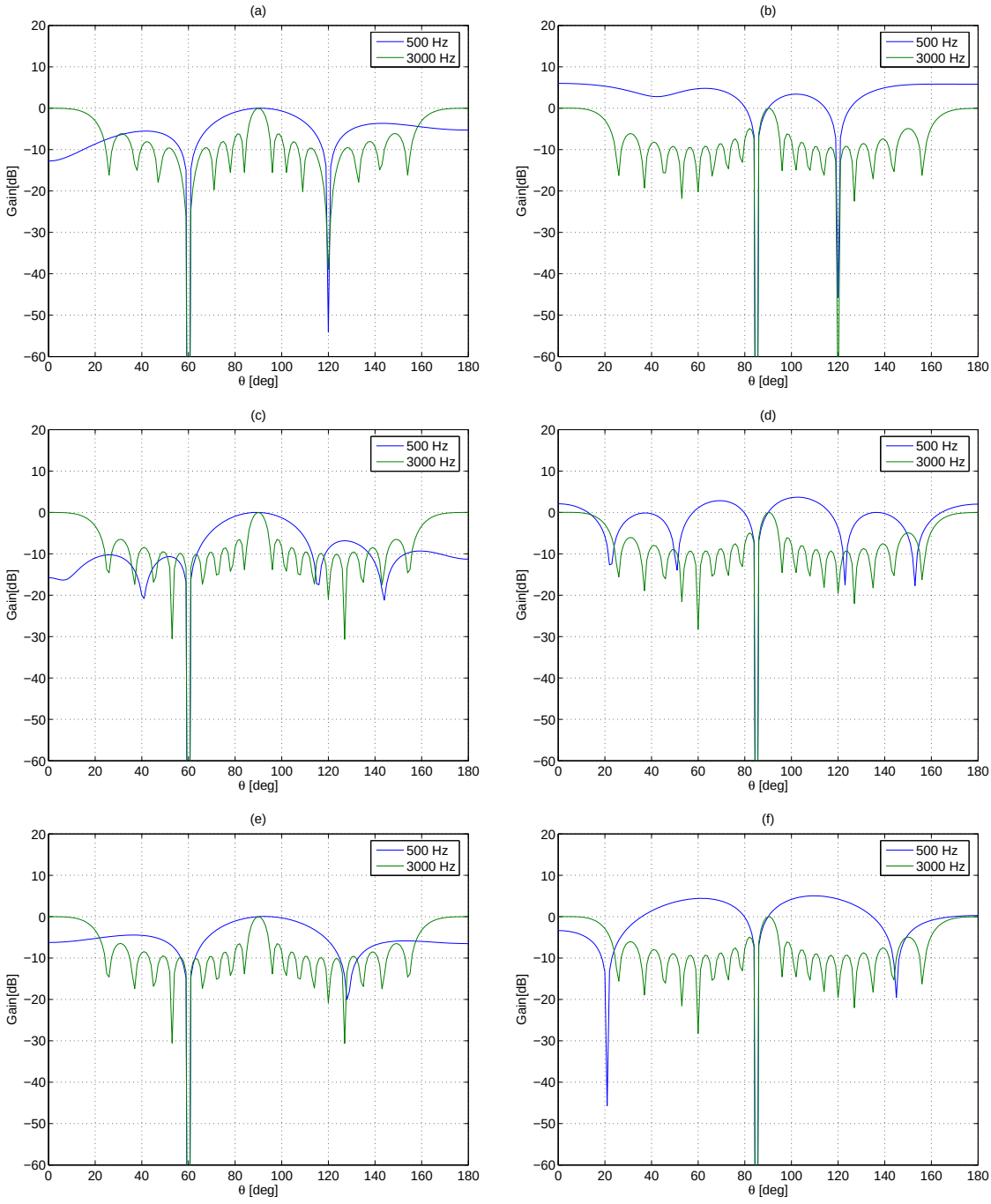


Figure 3.4: Spatial beamformer response in 500Hz and 3000Hz for the remote scenario: (a) Directional (c) Diffused, (e) Incoherent noise field; close scenario: (b) Directional (d) Diffused, (f) Incoherent noise field.

### Diffused Noise Field

If the array is designed to work with a diffused noise field, the deviation imposed on the desired signal is given in Figure 3.3 (c) and (d). Again, for the remote sources scenario, no more than 5dB deviation is imposed in the range $\theta \in [87, 93]$ for frequencies lower than 3000Hz. Similar results to the directional noise field are observed for the close sources scenario.

Note that for both scenarios the deviation is non symmetric as a function of the steering error, despite of the inherent symmetry of the noise field. This phenomenon is emphasized in the close sources scenario. This is due to the null directed towards the interference direction in the beamformer spatial response, implemented by both the MBF and BM. When the sources are remote, the interference reduction constraint has no affect on the deviation near the desired direction. On the other hand, when the sources are close, the contradicting constraints are hardly met causing a severe degradation around the desired direction.

The spatial beamformer response in 500Hz and 3000Hz is depicted in Figure 3.4 (c) and (d), for both the remote and close scenarios. Similar trends to the directional noise field can be observed.

### Incoherent Noise field

When the array is designed to work with an incoherent noise field, the deviation imposed on the desired signal is given in Figure 3.3 (e) and (f). Applying the remote sources scenario, no more than 5dB deviation is imposed in the range $\theta \in [87, 93]$ for frequencies lower than 3000Hz. On the other hand, in the close scenario, the deviation might reach -8dB, where maximum value of 0dB at $\theta = 90^o$ is obtained only in high frequencies. In fact, for low frequencies the maximum deviation is obtained in a direction which is outside the plot axis.

The spatial beamformer response in 500Hz and 3000Hz is depicted in Figure 3.4 (e) and (f). Similar trends to the directional noise field can be observed.

### 3.4.2 Signal's TFs: Real Room

For the evaluation of the general TFs case, we use ATFs simulated using Allen and Berkley's *image method* [26]. The scenario shown in Figure 3.5 is studied. The enclosure is a room with dimensions $4.5m \times 3.3m \times 4.2m$. A ten microphones array, with inter-element distance of $10cm$, is used. The sampling frequency is 8KHz, while reverberation time set to $T_{60} = 40$ms. The impulse response and the respective frequency response between the desired speech source and the first microphone are depicted in Figure 3.6.

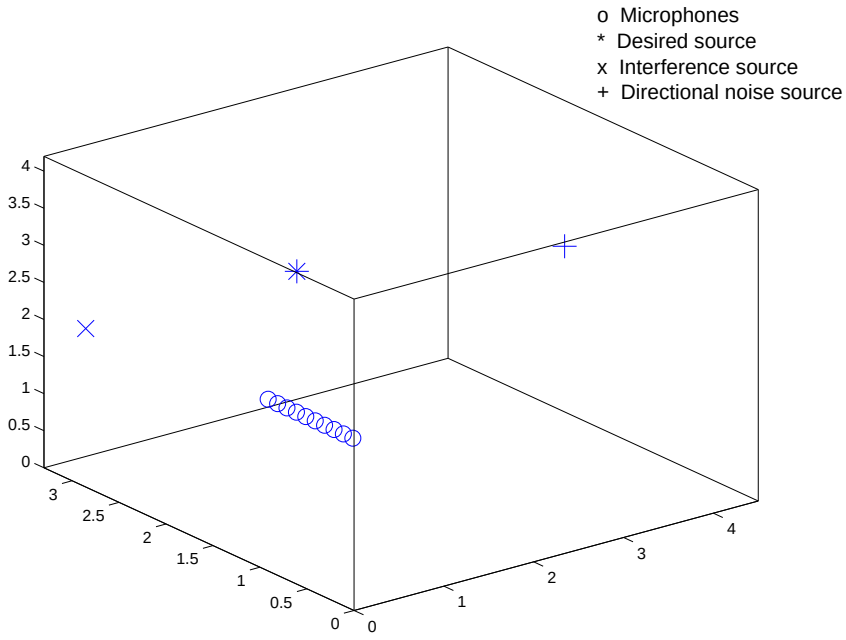


Figure 3.5: Test scenario: Array of ten microphones in a noisy and reverberated room.

We analyze the directional noise field case alone, since the deviation performance has no direct dependence on the noise type. Nevertheless, the noise field type may have minor influence on the estimation error and therefore affects the deviation. The algorithm was designed (i.e., $\boldsymbol{W_0}(e^{j\omega})$, $\mathcal{H}(e^{j\omega})$ and $\boldsymbol{G}(t, e^{j\omega})$ were determined) to receive signals from the desired direction, reject interference from the competing source direction and to cancel signals arriving from the noise di-

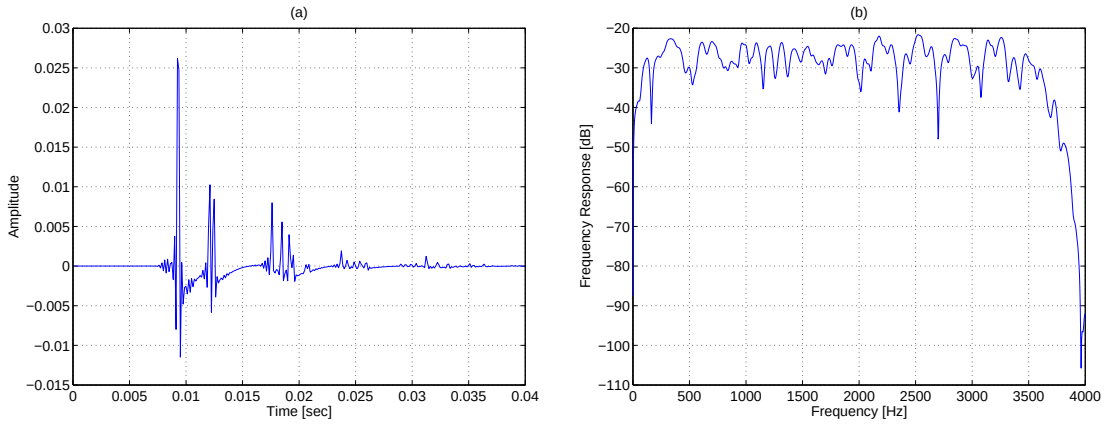


Figure 3.6: Ten microphones array in a noisy room: (a) Simulated impulse response (b) and frequency response of the ATF relates the desired speech signal and the first microphone.

rection. Results for the figure of merit introduced before (desired signal deviation $DEV(e^{j\omega})$) are presented. Note that the deviation is measured here twice. First, at the MBF output (denoted $DEV_{\mathrm{MBF}}(e^{j\omega})$), and second at the beamformer output (i.e. the total deviation, denoted $DEV(e^{j\omega})$). In Figure 3.7 the array response to a signal source transmitted from the desired direction is introduced. The deviation from the desired response demonstrated in the interesting frequency band is negligible. It can be noticed that the MBF succeeds in maintaining the desired signal direction, while the lower branch has almost no effect on the output.

Consider the case where, due to movement of the desired source or error in estimation of the ATFs ratios, the estimated desired source point is 35cm away off the real point (20cm away off in each axis). MBF deviation and array output deviation for that case are depicted in Figure 3.8 (a) and (b), respectively. It can be noticed that the deviation is no longer negligible. Results are presented in (c)-(f) for the cases in which the estimated point is 70cm and 105cm away off the real desired source. Clear degradation in performance can be observed as the distance increase. While 4dB average deviation can be observed in (b), 10dB average deviation is depicted in (d) and (f).

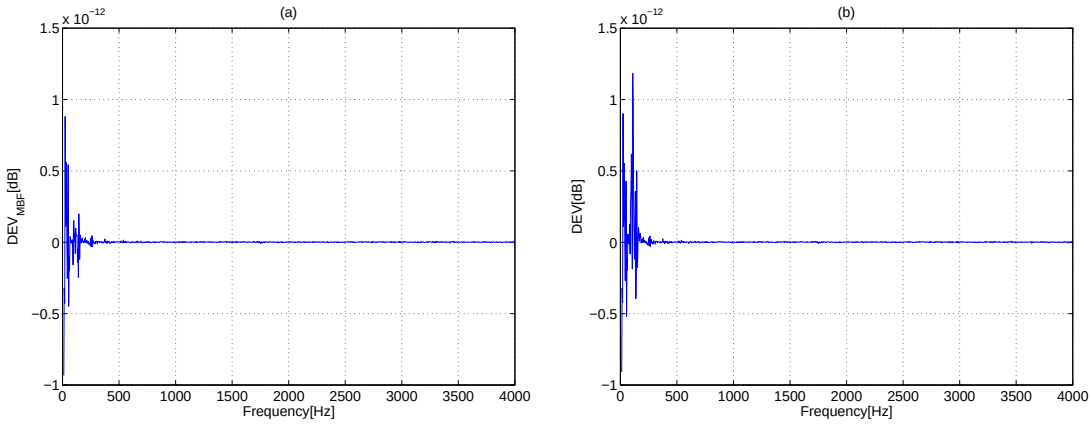


Figure 3.7: Expected Deviation performance for directional noise field: signal received from the desired direction. (a) MBF output deviation, (b) array output deviation.

Recall that the deviation plots depicted in Figure 3.8 were obtained for signals arriving from near the desired source. Since the BM blocks any signal which impinges on the array from the desired direction and its vicinity, considerable part of the received signals is blocked. In that case, the ANC uses low level reference signals $\boldsymbol{U}(t, e^{j\omega})$ and can't adapt properly. Therefore, it can be noticed that in all tested cases the lower branch has almost no effect on the deviation.

## 3.5 Evaluation of Noise Reduction

To calculate the noise reduction of the algorithm, we will use again the general expression for the output signal given by (3.7), this time with a noise signal $\boldsymbol{Z}(t, e^{j\omega}) = \boldsymbol{N}(t, e^{j\omega})$ (the same noise signal used to calculate the optimal Wiener filter) as the input signal. The expression (3.7) reduces to

$$\begin{aligned}\Phi^{n}_{yy}(t, e^{j\omega}) = &\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}) \\ &-\widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\left(\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1} \\ &\times\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}). \qquad (3.10)\end{aligned}$$

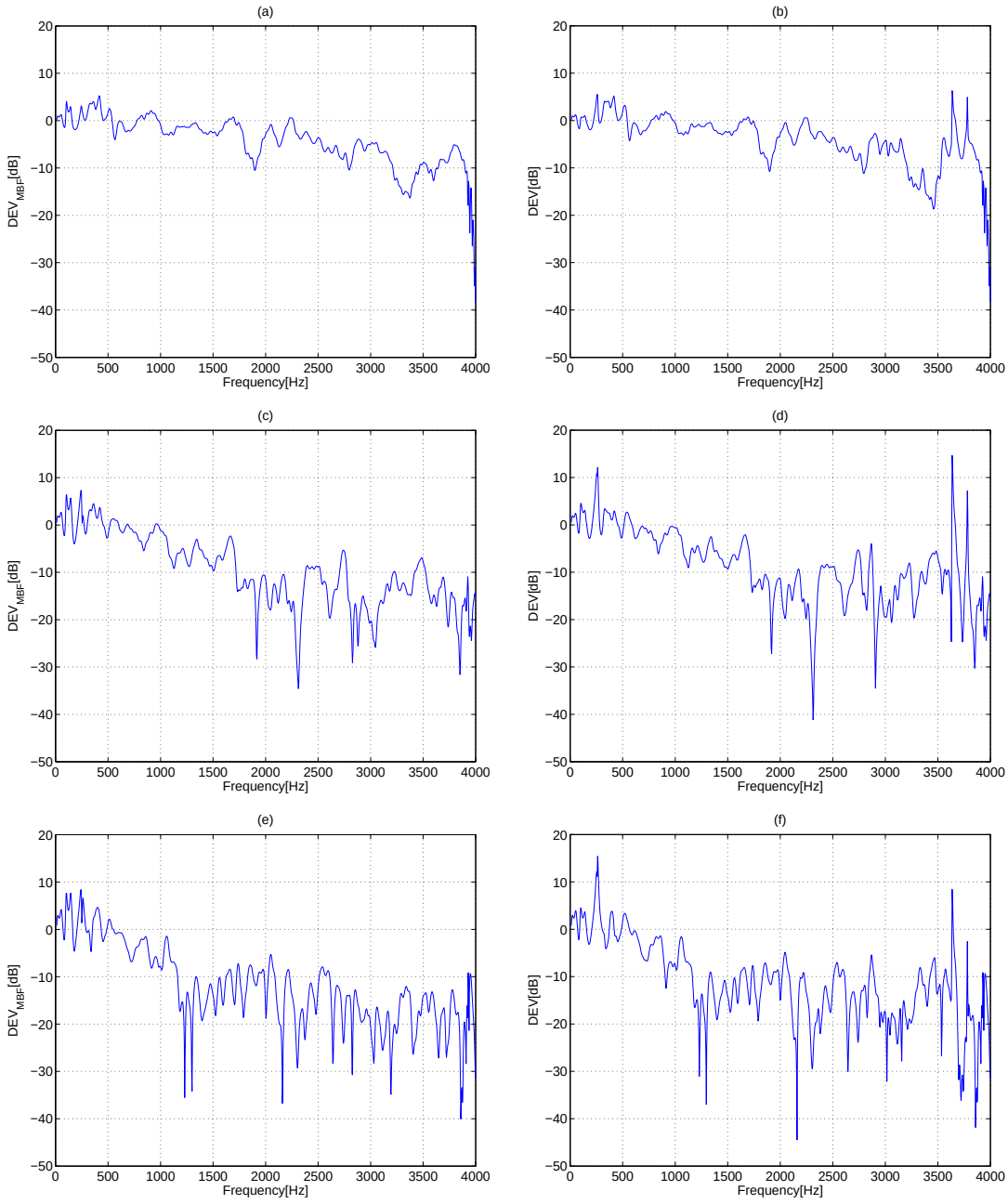


Figure 3.8: Expected deviation performance for directional noise field: desired signal received 35cm away off the estimated point: (a) MBF output deviation, (b) array output deviation; 70cm away off: (c) MBF output deviation, (d) array output deviation; 105cm away off: (e) MBF output deviation, (f) array output deviation.

Identify the first term of the equation as $\Phi^n_{\mathrm{MBF}}(t, e^{j\omega})$, the MBF part of the output PSD when only noise signal is present,

$$\begin{aligned}\Phi^n_{\mathrm{MBF}}(t, e^{j\omega}) &= E\{Y^n_{\mathrm{MBF}}(t, e^{j\omega}) \left(Y^n_{\mathrm{MBF}}(t, e^{j\omega})\right)^*\} \\ &= \widehat{\boldsymbol{W_0}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\widehat{\boldsymbol{W_0}}(e^{j\omega}). \end{aligned} \tag{3.11}$$

Expressions (3.10) and (3.11) can be used for calculating the noise reduction obtained by the algorithm and to determine the major contributor for this noise reduction.

The resulting expression for the noise cancellation depends on the noise PSD at the sensors. We calculate the expected noise reduction of the algorithm for three important noise fields: coherent (point source), diffused (spatially extended) and incoherent (noise signals generated at the sensors, e.g. amplifier noise, are assumed to be uncorrelated).

To evaluate the expected noise reduction of the algorithm, as we evaluated the predicted deviation, we should determine the TFs involved and the noise field. We begin with the simple steering array, where only delays relate the sources and the sensors. Thus, the DoA of the sources determines completely the TFs. The more general case where arbitrary TFs relate the sources and the sensors is more complicated to introduce, as the expected performance depends on the actual TFs used. For the evaluation of the general TFs case, we used the ATFs introduced in Section 3.4.2. We analyze now the dependence on the noise field.

### 3.5.1 Coherent Noise Field

Assume a single point source noise signal with PSD $\Phi_{nn}(t, e^{j\omega})$ and that $d_m(t)$ are slowly time varying transfer functions relating the noise source and the $m$-th sensor. Define,

$$\boldsymbol{N}(t, e^{j\omega}) = \boldsymbol{D}(e^{j\omega})N(t, e^{j\omega})$$

where,

$$\boldsymbol{D}(e^{j\omega}) = \left[\, D_1(e^{j\omega})\ D_2(e^{j\omega})\ \cdots\ D_M(e^{j\omega}) \,\right]^T.$$

The PSD matrix of the noise component of the sensors signals is given by,

$$\Phi_{\boldsymbol{NN}}(t, e^{j\omega}) = \Phi_{nn}(t, e^{j\omega})\boldsymbol{D}(e^{j\omega})\boldsymbol{D}^{\dagger}(e^{j\omega}) + \varepsilon I$$

where $I$ is an $M \times M$ identity matrix, and $\varepsilon \to 0$. The last term is added for stability reasons (see Appendix C). For $\boldsymbol{D}(e^{j\omega}) \neq \boldsymbol{A}(e^{j\omega})$, the achievable noise reduction is infinite, i.e.

$$\Phi_{yy}^{n}(t, e^{j\omega}) = 0 \;\text{ for }\; \boldsymbol{D}(e^{j\omega}) \neq \boldsymbol{A}(e^{j\omega}).$$

Thus, perfect noise cancellation is achieved, regardless the estimation accuracy. The derivation of this result is given in Appendix C. Note, that this is not a surprising result, since the presence of at least single reference signal for the noise ($M \geq 3$) is enough for the Wiener filter to completely eliminate the noise component. This result is valid for all TFs $\boldsymbol{D}(e^{j\omega})$ except for the signal direction TFs $\boldsymbol{D}(e^{j\omega}) = \boldsymbol{A}(e^{j\omega})$. When the desired source and noise source are located at the same point, namely $\boldsymbol{D}(e^{j\omega}) = \boldsymbol{A}(e^{j\omega})$, the noise and desired signal are indistinguishable and no noise reduction is expected. If $\boldsymbol{D}(e^{j\omega}) \neq \boldsymbol{A}(e^{j\omega})$ the suggested algorithm can eliminate any point source noise signal as good as the simple GSC can eliminate a directional noise signal in the delay-only propagation case (see for instance [29]), in addition to interference rejection inherent to the structure. It is also interesting to evaluate the noise part of the MBF branch, which is obtained due to the constraint

$$\begin{aligned}
\Phi_{\mathrm{MBF}}^{n}(t, e^{j\omega}) &= \boldsymbol{W_0}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\boldsymbol{W_0}(e^{j\omega}) \\
&= \boldsymbol{W_0}^{\dagger}(e^{j\omega})\Phi_{nn}(t, e^{j\omega})\boldsymbol{D}(e^{j\omega})\boldsymbol{D}^{\dagger}(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}) \\
&= \Phi_{nn}(t, e^{j\omega})|\boldsymbol{W_0}^{\dagger}(e^{j\omega})\boldsymbol{D}(e^{j\omega})|^2. \qquad (3.12)
\end{aligned}$$

The $|\boldsymbol{W_0}^{\dagger}(e^{j\omega})\boldsymbol{D}(e^{j\omega})|^2$ factor depends on the estimated MBF and the ATF $\boldsymbol{D}(e^{j\omega})$ and might be greater than one. Hence, the MBF does not necessarily improves the desired signal to noise ratio, even though it combines the desired signal coherently. The infinite noise reduction is due to the noise canceller branch.

**Signal's TFs: Pure Delay**

Assume free space propagation, i.e., the ATFs from the desired source and interference source to the sensors are pure delay. Signal is assumed to impinge on the array from $\theta = 90^o$, while the inter-element distance is 10cm. As in the previous section, we will examine two scenarios. In the first, the nonstationary interference impinges on the array from $\theta = 60^o$ (denoted as remote scenario), while in the second, $\theta = 85^o$ (denoted as close scenario).

We optimize the array to cancel noise source from $\theta = 120^o$. In Figure 3.11 (a) and (b) we present the PSD of the array output as a function of the frequency and the array steering angle $\theta$. For both scenarios, it is clearly shown that the main lobe is maintained, while null is constructed in all frequencies at the noise angle (in addition to null at the interference angle).

Recall the spatial beamformer response in 500Hz and 3000Hz was depicted in Figure 3.4 (a) and (b), for the remote and close scenarios, respectively. It can be clearly seen that the main lobe around the desired direction is wider in 500Hz than in 3000Hz. When the directional noise source is away from the desired source, this phenomenon is meaningless. Nevertheless, when the noise arrives from a direction adjacent to the desired source, the BM, constrained to block this direction, prevents the ANC from cancelling the unwanted components. Furthermore, it was shown that for low frequencies the maximum deviation is obtained in a direction which is outside the plot axis. Therefore, the MBF might enhance low frequency noise signal arriving from a direction adjacent to desired source, which could result in large noise component to the beamformer output.

**Signal's TFs: Real Room**

The algorithm was designed (i.e., $\boldsymbol{W_0}(e^{j\omega})$, $\mathcal{H}(e^{j\omega})$ and $\boldsymbol{G}(t, e^{j\omega})$ were determined) to receive signals from the desired direction, reject interference from the competing source direction and to cancel signals arriving from the noise direction. Note that noise reduction might be obtained due to the acoustic environment, especially when the noise source is far from the microphones array. Therefore, in

order to eliminate the acoustic environment contribution to the noise reduction, the NR is calculated by normalizing the output PSD by the ATF that relates the noise source and the first microphone. We assume that since the microphones are close one to each other, the ATFs which relate the noise source and the microphones, are similar, and averaged normalization can be obtained by using one of the ATFs.

A value of $\mathrm{NR}(t, e^{j\omega}) = 0$ indicates complete noise reduction. Note that $NR(e^{j\omega})$ is measured here twice. First, at the MBF output (denoted $NR_{\mathrm{MBF}}(e^{j\omega})$), and second at the beamformer output (denoted $NR(e^{j\omega})$). Results for the two figures of merit introduced in Figure 3.9, for a signal source transmitted from the noise direction.

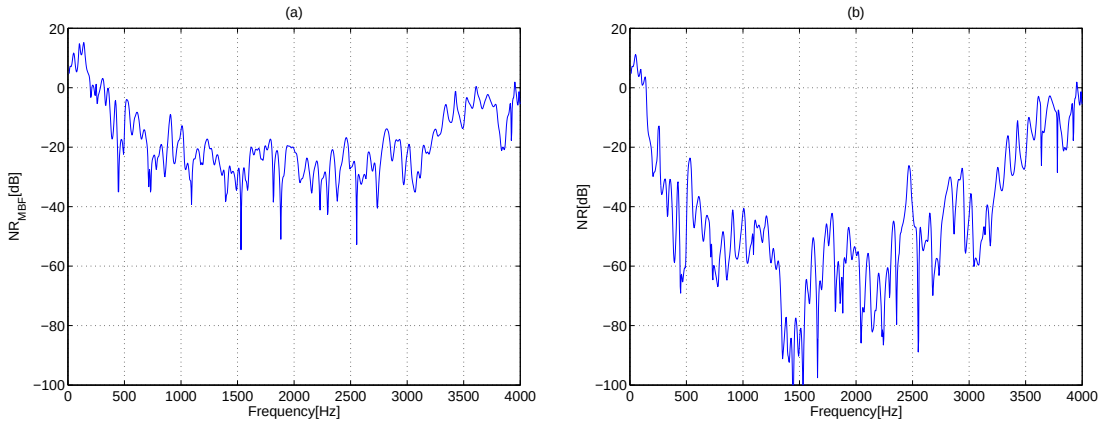


Figure 3.9: Expected NR performance for directional noise field: signal received from the noise direction. (a) MBF output PSD, (b) array output PSD.

Noise reduction of more than 30 dB is demonstrated in most of the interesting frequency band. It can be noticed that part of the noise reduction is due to the MBF and part is due to the lower branch. Some amplification can be observed at frequencies below 200Hz and performance degradation can be noticed above 3300Hz. Nevertheless, it should be emphasized that speech signal usually contains low energy in these bands.

For the evaluation of NR in the presence of estimation errors, a beamformer

was designed to reject noise signal according to the system introduced in Section 3.4.2, while the noise is actually arriving from a different source. This evaluation was conducted for three different distances between the assumed noise source and real point: 17cm, 35cm and 70cm. The NR degradation is calculated by subtracting the NR achieved using exact ATFs, from the NR obtained in the cases above.

Degradation in NR performance for distance of 17cm is depicted in Figure 3.10 (a) and (b), as measured at the MBF output and beamformer output, respectively. While no clear trend can be observed in the MBF performance, 25dB average degradation in performance can be observed in the beamformer output. Hence, most of the degradation is due to the inaccurate ANC. The degradation in performance increases as the error in the estimation steps up, as depicted in Figure 3.10 (c) and (d), measured for 35cm estimation error. Average degradation of 40dB is measured at the beamformer output. Similar results are obtained if the noise source is transmitted 70cm away off the assumed point, as depicted in Figure 3.10 (e) and (f).

### 3.5.2 Incoherent Noise Field

For incoherent noise field we assume that the noise at the sensors has no spatial correlation. Thus,

$$\Phi_{\boldsymbol{NN}}(t, e^{j\omega}) = \Phi_{nn}(t, e^{j\omega})I,$$

where $I$ is an $M \times M$ identity matrix. Using (3.10) and the prespecified $\Phi_{\boldsymbol{NN}}(t, e^{j\omega})$ we obtain,

$$\Phi^n_{yy}(t, e^{j\omega}) = \Phi_{nn}(t, e^{j\omega})\boldsymbol{W_0}^{\dagger}(e^{j\omega}) \left\{ I - \mathcal{H}(e^{j\omega}) \left( \mathcal{H}^{\dagger}(e^{j\omega})\mathcal{H}(e^{j\omega}) \right)^{-1} \mathcal{H}^{\dagger}(e^{j\omega}) \right\} \boldsymbol{W_0}(e^{j\omega}).$$

Using (2.37) ($\widehat{\boldsymbol{A}}^{\dagger}(e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) = 0$ and $\widehat{\boldsymbol{B}}^{\dagger}(e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) = 0$) we observe that

$$\widehat{\boldsymbol{W}_0}^{\dagger}(e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega}) =$$

$$\left[ \frac{\left\| \frac{\widehat{\boldsymbol{B}(e^{j\omega})}}{B_1(e^{j\omega})} \right\|^2 \frac{\widehat{\boldsymbol{A}(e^{j\omega})}}{A_1(e^{j\omega})} - \frac{\widehat{\boldsymbol{B}(e^{j\omega})}}{B_1(e^{j\omega})} \frac{\widehat{\boldsymbol{B}^{\dagger}(e^{j\omega})}}{B_1^*(e^{j\omega})} \frac{\widehat{\boldsymbol{A}(e^{j\omega})}}{A_1(e^{j\omega})}}{\left\| \frac{\widehat{\boldsymbol{A}(e^{j\omega})}}{A_1(e^{j\omega})} \right\|^2 \left\| \frac{\widehat{\boldsymbol{B}(e^{j\omega})}}{B_1(e^{j\omega})} \right\|^2 - \frac{\widehat{\boldsymbol{A}^{\dagger}(e^{j\omega})}}{A_1^*(e^{j\omega})} \frac{\widehat{\boldsymbol{B}(e^{j\omega})}}{B_1(e^{j\omega})} \frac{\widehat{\boldsymbol{B}^{\dagger}(e^{j\omega})}}{B_1^*(e^{j\omega})} \frac{\widehat{\boldsymbol{A}(e^{j\omega})}}{A_1(e^{j\omega})}} \mathcal{F}(e^{j\omega}) \right]^{\dagger} \widehat{\mathcal{H}}(e^{j\omega}) = \mathbf{0}_{1\times(M-2)}.$$

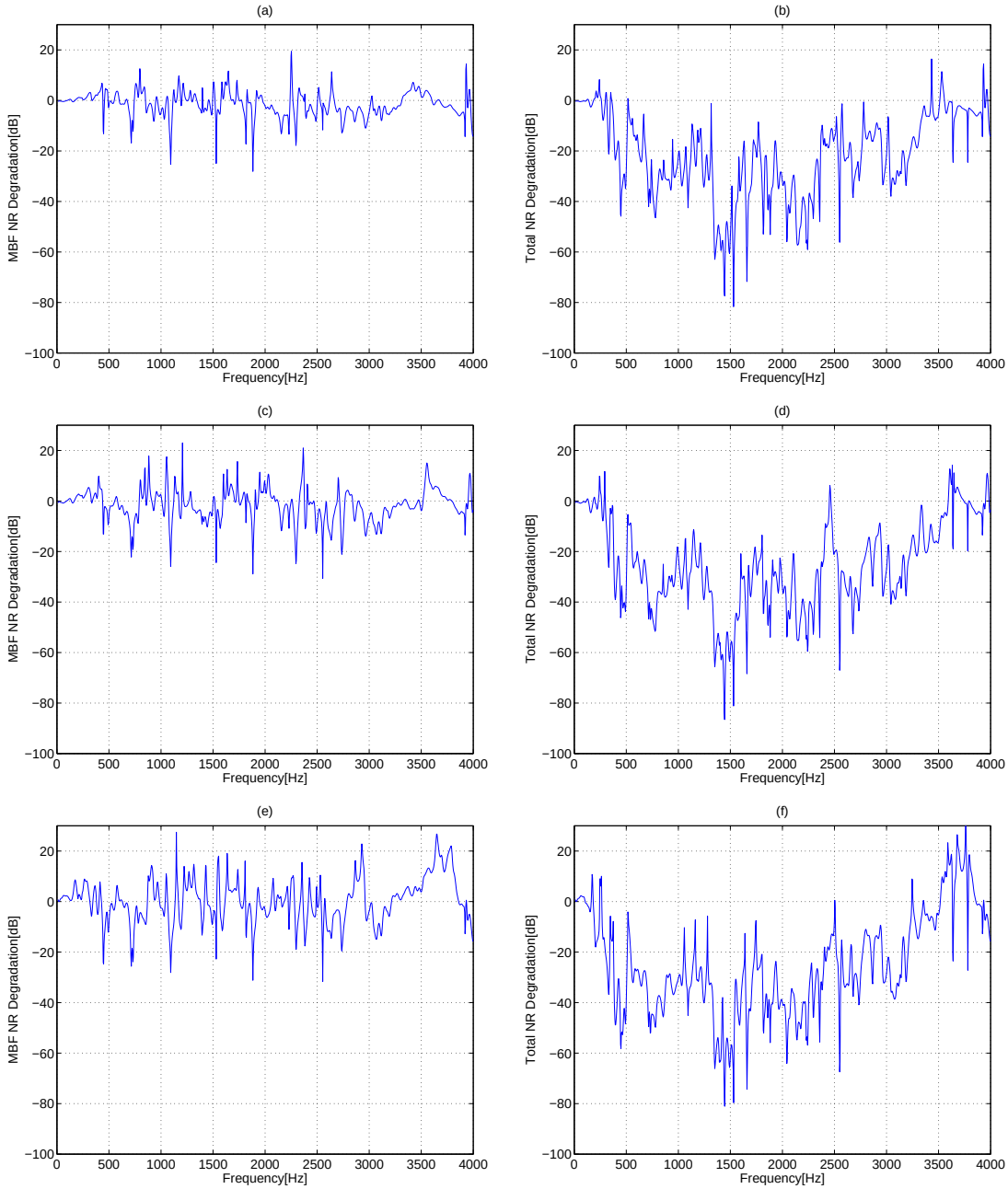


Figure 3.10: Expected NR degradation for directional noise field: signal received from 17cm near the noise source: (a) MBF NR degradation, (b) output NR degradation; 35cm near the noise source: (c) MBF NR degradation, (d) output NR degradation; 70cm near the noise source: (e) MBF NR degradation, (f) output NR degradation.

Note that the last transition is valid regardless of the estimation accuracy, when both $\widehat{\boldsymbol{W}_0}(e^{j\omega})$ and $\widehat{\mathcal{H}}(e^{j\omega})$ are calculated using the same estimated $\frac{\widehat{\boldsymbol{A}(e^{j\omega})}}{A_1(e^{j\omega})}$ and $\frac{\widehat{\boldsymbol{B}(e^{j\omega})}}{B_1(e^{j\omega})}$. When $\widehat{Q_m}(e^{j\omega})$ and $\widehat{L_m}(e^{j\omega})$ are estimated directly, this might not be guaranteed.

Furthermore, as $\mathcal{H}^{\dagger}(e^{j\omega})\mathcal{H}(e^{j\omega})$ is a positive matrix, its inverse always exists. Thus, the contribution of the noise cancelling branch is zero, and the noise reduction is only due to the matched beamformer. The noise power at the output is thus,

$$\Phi_{yy}^{n}(t, e^{j\omega}) = \Phi_{\text{MBF}}^{n}(t, e^{j\omega}) = \Phi_{nn}(t, e^{j\omega})\boldsymbol{W_0}^{\dagger}(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}).$$

Again, no noise reduction is guaranteed by this structure, and the result depends on the TFs' ratio involved.

#### Signal's TFs: Pure delay

Array output PSD is depicted in Figure 3.11 (e) and (f) for various steering angles and the entire frequency band. Similar to the diffused noise case, an LCMV beamformer in the presence of incoherent noise, maintains the desired direction while mitigating all others. Less than -10db white noise gain can be observed for most of the interesting frequency band.

#### Signal's TFs: Real Room

MBF output and array output PSD for incoherent noise field are depicted in Figure 3.12. Since only minor difference between the two plots can be observed, it can be assumed that no noise reduction is achieved by the noise cancelling branch, and the negligible noise reduction is due to the MBF branch only.

### 3.5.3 Diffused Noise Field

In highly reverberant acoustical environment, such as a car enclosure, the noise field tends to be diffused (see for instance [27] and [23]). A diffused noise source is

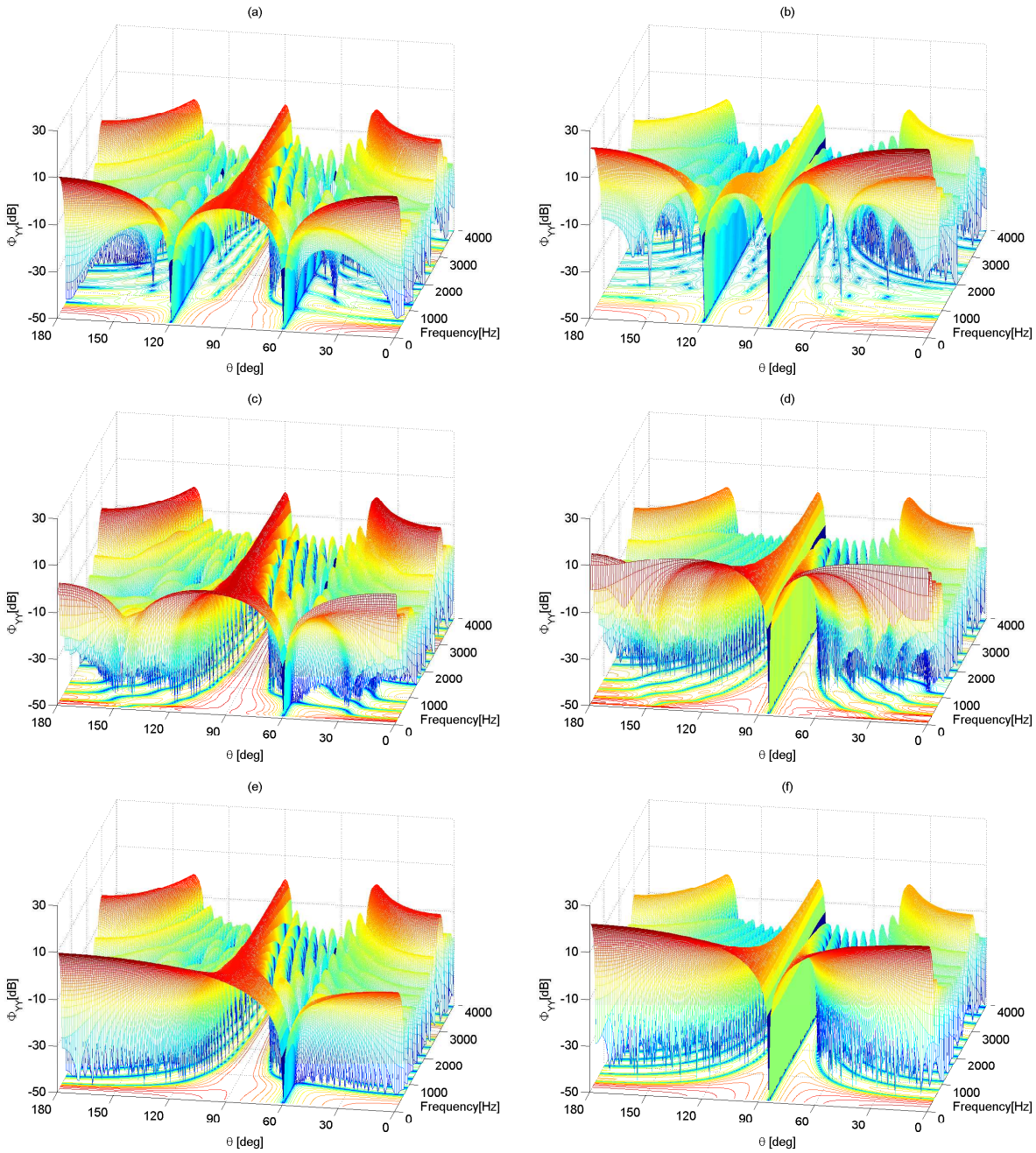


Figure 3.11: Array output PSD $\Phi_{yy}$ as a function of the frequency and direction of arrival. Desired signal direction $\theta = 90^o$. $M = 10$ sensors. remote sources: (a) Directional noise field ($\theta = 120^o$), (c) Diffused noise field, (e) Incoherent noise field; close sources: (b) Directional noise field ($\theta = 120^o$), (d) Diffused noise field, (f) Incoherent noise field.

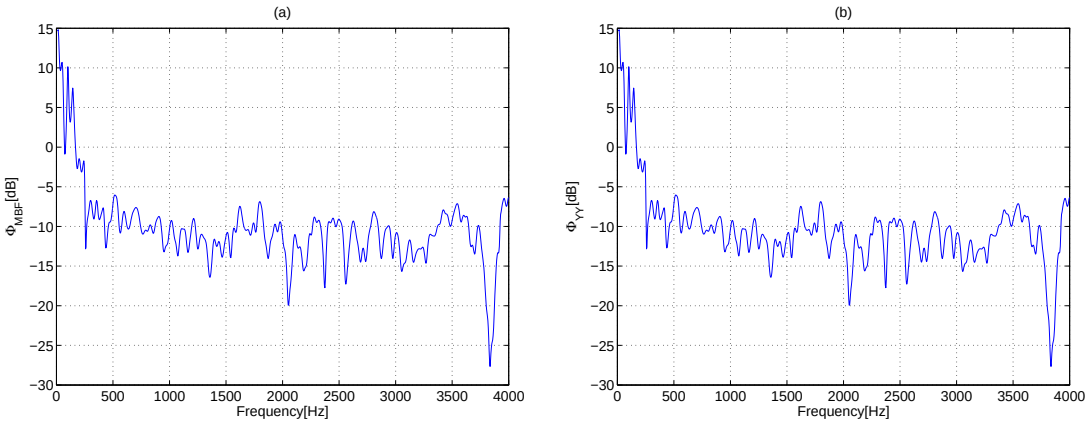


Figure 3.12: Expected NR performance for incoherent noise field: (a) MBF output PSD, (b) array output PSD.

assumed to be equidistributed on a sphere in the far field of the array. The cross-coherence function between signals received by two sensors $(i,j)$ with distance $d_{ij}$ can be found in [27], and is given in the following expression:

$$\Gamma_{N_iN_j}(e^{j\omega}) = \frac{\Phi_{N_iN_j}(e^{j\omega})}{\sqrt{\Phi_{N_iN_i}(e^{j\omega})\Phi_{N_jN_j}(e^{j\omega})}} = \frac{\sin(\omega d_{ij}/c)}{\omega d_{ij}/c},$$

where $c$ is the speed of sound. Therefore, the coherence matrix is given by,

$$\Gamma(e^{j\omega}) = \begin{bmatrix} 1 & \Gamma_{N_1N_2}(e^{j\omega}) & \cdots & \Gamma_{N_1N_M}(e^{j\omega}) \\ \Gamma_{N_2N_1}(e^{j\omega}) & 1 & \cdots & \\ & & \ddots & \\ \Gamma_{N_MN_1}(e^{j\omega}) & & & 1 \end{bmatrix}.$$

The noise PSD at the sensors input is thus,

$$\Phi_{\boldsymbol{NN}}(t, e^{j\omega}) = \Phi_{nn}(t, e^{j\omega})\Gamma(e^{j\omega}).$$

Using (3.11) the noise PSD at the MBF output is given by,

$$\Phi^n_{\mathrm{MBF}}(t, e^{j\omega}) = \Phi_{nn}(t, e^{j\omega})\boldsymbol{W_0}^{\dagger}(e^{j\omega})\Gamma(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}),$$

and using (3.10) the noise PSD at the beamformer output is obtained

$$\begin{aligned}\Phi^n_{yy}(t, e^{j\omega}) &= \boldsymbol{W_0}^{\dagger}(e^{j\omega})\Gamma(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}) \\ &-\boldsymbol{W_0}^{\dagger}(e^{j\omega})\Gamma(e^{j\omega})\mathcal{H}(e^{j\omega})\left(\mathcal{H}^{\dagger}(e^{j\omega})\Gamma(e^{j\omega})\mathcal{H}(e^{j\omega})\right)^{-1}\mathcal{H}^{\dagger}(e^{j\omega})\Gamma(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}).\end{aligned} \tag{3.13}$$

This expression depends on the TFs ratios, assumed to be error free, and the coherence function $\Gamma(e^{j\omega})$.

Note that in frequencies above the cutoff frequency $\frac{c}{2d}$, where $d$ is the inter element distance, the elements of $\Gamma(e^{j\omega})$ are almost zero besides the main diagonal elements. In this case, $\Phi^n_{yy}(t, e^{j\omega})$ is close to the incoherent noise case. On the other hand, in lower frequencies, $\Gamma(e^{j\omega})$ components are almost one. In that case, $\Phi^n_{yy}(t, e^{j\omega})$ is close to the coherent noise case. In our simulations, the inter-element distance between the microphones is $10cm$, and therefore the cutoff frequency is 1500Hz.

**Signal's TFs: Pure delay**

The array was analyzed with diffused noise field by using (3.13). The case of $M = 10$ sensors is shown in Figure 3.11 (c) and (d) for various steering angles and the entire frequency band. Recall that the diffused noise impinges on the array from all directions. Therefore, the beamformer attenuates almost all directions, except the desired one. More than 10db attenuation can be observed for most frequencies in the interesting frequency band. It can be observed that in low frequencies, low attenuation is obtained in the remote scenario and even some amplification in the close scenario. Note that when considering noise reduction in the presence of diffused noise, wide main lobe as seen in Figure 3.4 (c) means that more noise would leak into the beamformer output.

**Signal's TFs: Real Room**

The MBF output PSD and the beamformer output PSD are given in Figure 3.13 for the diffused noise field. It is clearly shown, that the expected performance of the algorithm in diffused field is inferior to the expected performance for point source noise field. This is verified by the experimental study of Chapter 2. Most of the noise reduction is achieved by the MBF. Note that the MBF amplifies frequencies below 300Hz, while some noise reduction is obtained by the lower branch in frequencies below 500Hz. Nevertheless, it should be emphasized that a

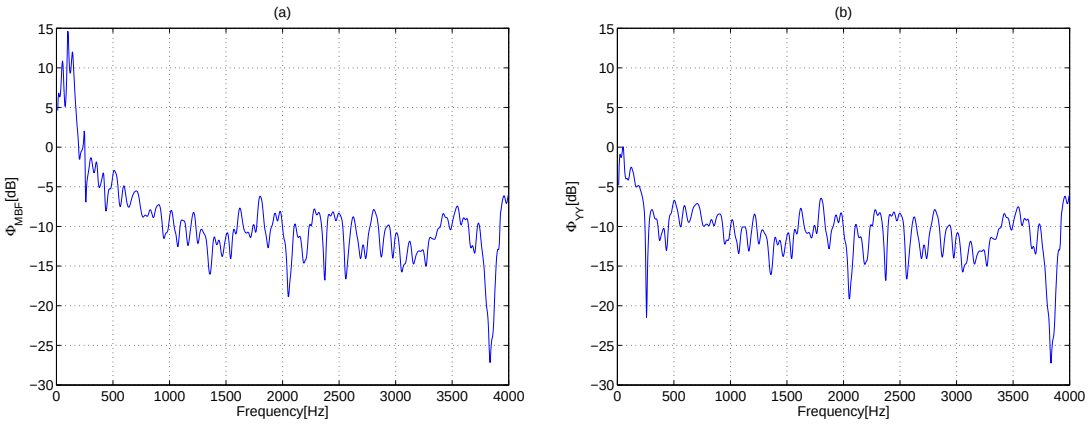


Figure 3.13: Expected NR performance for diffused noise field: (a) MBF output PSD, (b) array output PSD.

speech signal usually contains low energy in this band. Hence, the lower branch renders useless for most of the interesting band in this case.

## 3.6 Evaluation of Interference Reduction

The interference reduction of the algorithm can be calculated by the general expression given in (3.7) for a signal $\boldsymbol{Z}(t, e^{j\omega}) = \boldsymbol{B}(e^{j\omega})s_2(t, e^{j\omega})$. Assume we have exact knowledge of the TFs' ratio $\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}$ and $\frac{\boldsymbol{B}(e^{j\omega})}{B_1(e^{j\omega})}$, i.e. $\widehat{\mathcal{H}}(e^{j\omega}) = \mathcal{H}(e^{j\omega})$ and $\widehat{\boldsymbol{W_0}}(e^{j\omega}) = \boldsymbol{W_0}(e^{j\omega})$. Using $\boldsymbol{B}^{\dagger}(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}) = 0$ and $\mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{W_0}(e^{j\omega}) = 0$, $\Phi_{yy}(t, e^{j\omega}) = 0$ can be obtained. Note the above result does not depend on the noise field type. Furthermore, opposed to noise reduction, the interference reduction is due to the constraints imposed on the MBF and BM alone, rather than minimization or adaptive filtering.

On the other hand, when considering real scenario, the TFs ratios estimates are not accurate, and $\widehat{Q}(e^{j\omega})$ and $\widehat{L}(e^{j\omega})$ are not exact estimates of $Q(e^{j\omega})$ and $L(e^{j\omega})$, respectively. In that case, the noise field type might affect the estimation performance and therefore affect the interference blocking ability of the MBF and the quality of the noise reference at the BM output, as well. Both impaired MBF and presence of interference components in the noise reference signals, result in

interference leakage at the beamformer output.

We now evaluate the interference reduction performance in the presence of estimation errors. As with the deviation evaluation, we begin with simple steering array, where the DoA determines completely the TFs. Afterwards, we deal with real room ATFs.

### Signal's TFs: Pure Delay

Assume free space propagation, i.e., the ATFs from the desired source and interference source to the sensors are pure delay. Desired signal is assumed to impinge on the array from $\theta = 90^o$, while the inter-element distance is 10cm. We will examine two scenarios. In the first, denoted *remote sources*, the interference signal impinges on the array from $\theta = 60^o$, while in the second, denoted *close sources*, $\theta = 85^o$. We analyze now the dependence on the noise field.

The array output PSD in the presence of directional noise is depicted in Figure 3.14 (a) and (b), for both tested scenarios. In both, a 60dB null can be observed in the interference direction, $\theta = 60^o$ for the remote case and $\theta = 85^o$ for the close case.

The array output PSD in the presence of diffused noise is depicted in Figure 3.14 (c) and (d), for both tested scenarios. As for the directional noise, a 60dB null can be observed in the interference direction. Only minor changes can be noticed, compared to the directional noise field.

The array output PSD in the presence of incoherent noise is depicted in Figure 3.14 (e) and (f), for both tested scenarios. The same trends can be observed for the incoherent noise field.

### Signal's TFs: Real Room

For the evaluation of the general TFs case, we used the same ATFs introduced in Section 3.4.2. Similarly to noise reduction evaluation, a beamformer was designed to reject an interference signal according to the system introduced in Section 3.4.2, while the interference arrives from a different direction than assumed.

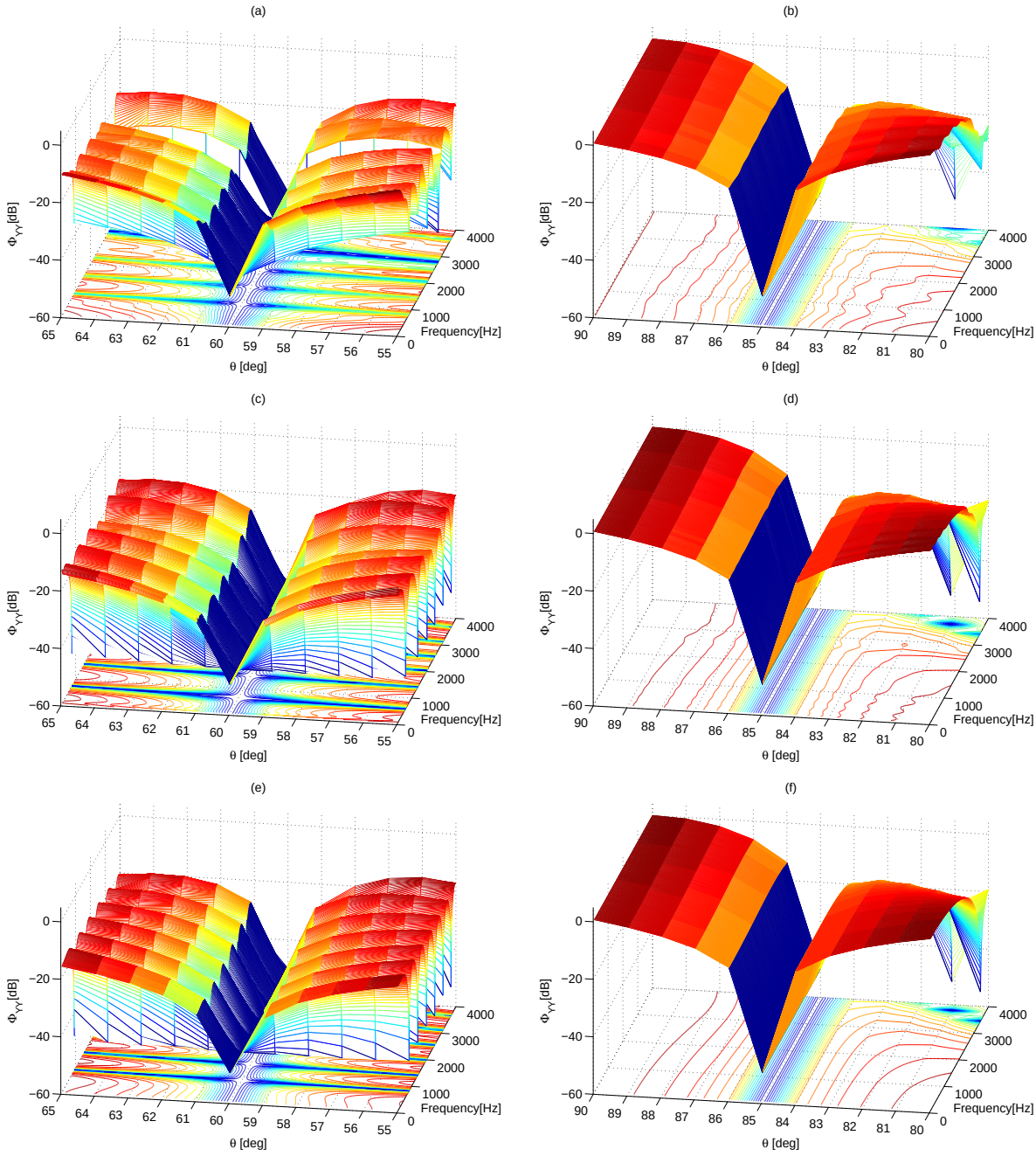


Figure 3.14: Array output PSD $\Phi_{yy}$ as a function of the frequency and direction of arrival. $M = 10$ sensors. Remote sources, interference signal direction $\theta = 60^o$: (a) Directional noise field ($\theta = 120^o$), (c) Diffused noise field, (e) Incoherent noise field; Close sources, interference signal direction $\theta = 85^o$: (b) Directional noise field ($\theta = 120^o$), (d) Diffused noise field, (f) Incoherent noise field.

This evaluation was conducted for three different distances between the assumed interference source and real point: 3.5cm, 35cm and 105cm.

Note that similar to noise signal, the interference signal might be attenuated due to the acoustic environment. Therefore, the interference reduction is calculated by normalizing the output PSD by the ATF that relates the interference source and the first microphone. We therefore define the normalized interference reduction by,

$$\mathrm{NIR}(t, e^{j\omega}) = \frac{\Phi_{yy}^{s_2}(t, e^{j\omega})}{|\mathcal{F}(e^{j\omega})|^2 |B_1(e^{j\omega})|^2 \Phi_{s_1 s_1}(t, e^{j\omega})}. \tag{3.14}$$

A value of $\mathrm{NIR}(t, e^{j\omega}) = 0$ indicates complete interference reduction. This value is obtained whenever an exact knowledge of the TFs ratios is available.

The results for the three cases are depicted in Figure 3.15. Note that $NIR(e^{j\omega})$ is measured here twice. First, at the MBF output (denoted $NIR_{\mathrm{MBF}}(e^{j\omega})$), and second at the beamformer output (denoted $NIR(e^{j\omega})$). The array was designed assuming directional noise field.

Note that in all three distances, the interference reduction is due to the MBF alone, since the BM blocks the interference and therefore the ANC does not affect the output. The average normalized interference reduction for the 3.5cm and 35cm cases is 30dB and 20dB, respectively. Some interference reduction can be measured even when the real interference source is more than 1m away from the assumed point. Nevertheless, in that case the beamformer emphasize some frequencies above 3500Hz.

## 3.7 Summary

Due to signal source movement, change in the acoustic environment or estimation errors, the DTF-GSC algorithm might use inaccurate ATFs. In this chapter we evaluated the degradation in performance due to these estimation errors. We derived a general expression for the output power signal density. From this expression we derived expressions for the expected deviation imposed on the desired

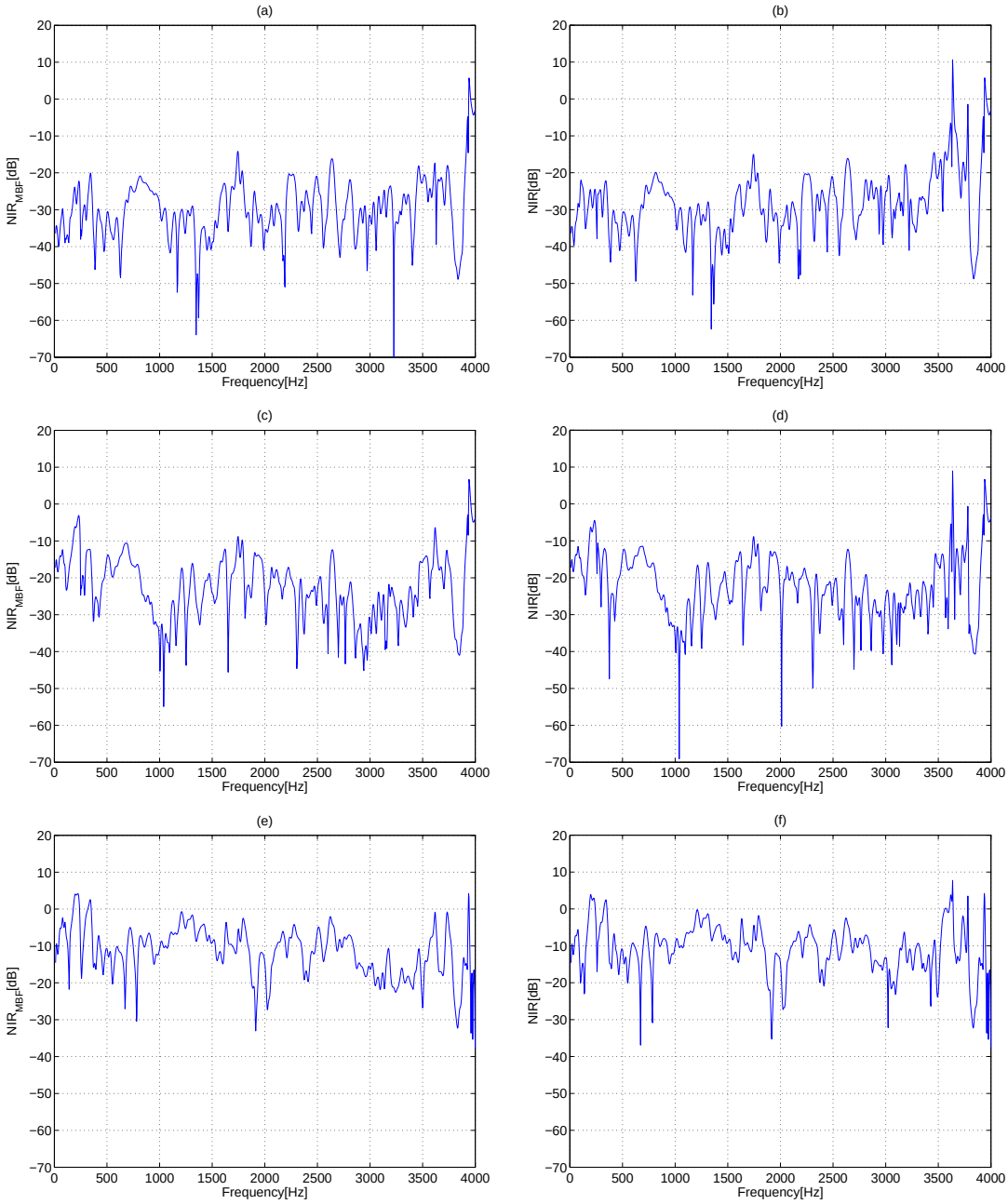


Figure 3.15: Expected normalized IR performance for directional noise field: 3.5cm near interference source (a) MBF output PSD, (b) array output PSD; 35cm near interference source (c) MBF output PSD, (d) array output PSD; 105cm near interference source (e) MBF output PSD, (f) array output PSD.

signal and the amount of achievable noise and interference reduction. Gannot *et al.* [28] presented an analysis of the PSD deviation imposed on the desired signal by the transfer function generalized sidelobe canceller (TF-GSC) algorithm. However, in this contribution we deal with the dual nonstationary interference problem (DTF-GSC) and present a noise reduction and interference reduction analysis as well.

The deviation and interference reduction depend on the quality of estimating the ratio of TFs. The noise reduction depends on the TFs' ratio involved as well as the noise field. High noise reduction is achieved in a coherent noise field. In an incoherent noise field only the MBF branch is responsible for noise reduction. The amount of noise reduction depends on the TFs' ratio involved, and it might be negligible. In a diffused noise field we expect a significant noise reduction only in the lower frequency band, but when there is almost no desired signal in that band, the amount of noise reduction may also be almost negligible. We note that in actual scenarios the noise field is a mixture between point sources and diffused sources, thus the expected result should fall between these two extremes. Recall that in practice, the suggested algorithm uses sequential update for the filters involved rather than the closed form optimal solution. This might degrade the performance of the algorithm, yielding results which are inferior to those presented in this study.

# Chapter 4

# Joint Noise Reduction and Acoustic Echo Cancellation

Man machine interaction requires acoustic interface in order to provide full duplex hands-free communication. In Chapter 2 we considered an environment where the desired signal is contaminated by both competing speech signal and noise, while no reference signal is available. Nevertheless, in many cases, such as hands-free communication, the competing signal is produced by the coupling of the loudspeaker and microphone, i.e. an echo signal. In that case, the echo signal can be exploited. In a noisy environment where desired signal and echo signal coexist, the tasks of echo cancellation and noise reduction contradict each other.

In this chapter, a DTF-GSC based scheme is suggested to jointly reduce noise and cancel echo. This scheme is evaluated through a series of experiments in single and double talk situations, and compared to two straight forward schemes, implemented in the frequency domain.

## 4.1 Background

In many speech communication applications, like audio-conference and hands-free mobile telephone, the received multi-microphone speech signals are corrupted by acoustic background noise as well as echo signals. This causes a signal degradation which can lead to speech unintelligibility and degradation in the performance of

subsequent speech coding devices and speech recognition systems.

Note that echo signals alone can be suppressed successfully by *acoustic echo canceller* (AEC) if no noise signal exists, while background noise can be reduced using fixed or adaptive beamformer (BF). In [14],[30] Kellermann decoupled the problem and proposed two generic joint schemes, multi-channel AEC followed by a beamformer (denoted AEC-BF) and beamformer followed by a single channel AEC (denoted BF-AEC). The blocks proposed in these schemes are implemented in the time domain. However, The time domain is known to be less applicable in real-life scenarios where complex ATFs relate the source and the microphone signals.

In [12], two frequency domain schemes for joint echo cancellation and noise reduction are presented. Both contain the TF-GSC beamformer proposed in [5] and a *block least mean square* (BLMS) AEC. Following Kellermann [14], the first scheme comprises multi-channel AEC followed by a beamformer, while the second comprises a beamformer followed by a single channel AEC as a post-filter. A series of simulations using real speech recordings showed that the first scheme outperforms the second one. However, it has been shown that both schemes suffer from disadvantages since the tasks of echo cancellation and noise reduction contradict each other. When the AEC precede the beamformer, the AEC performance impaired due to the existence of noise signals, while the performance of the AEC in the BF-AEC severely deteriorates due to the variation of the echo path caused by the beamformer.

Herbordt and Kellermann suggested in [31] the GSAEC scheme, an AEC embedded into GSC beamformer, and compared it to the previously presented AEC-BF and BF-AEC, when using the GSC as beamformer. In this structure, as depicted in Figure 4.1 ,the AEC module is placed in the upper branch of the GSC, behind the *fixed beamformer* (FBF). Therefore, only single AEC is required for an arbitrary number of array elements, opposed to the AEC-BF scheme. Experiments confirmed that the noise reduction is degraded by less than 3 dB compared to a cascade of AEC and GSC and that the echo reduction is

improved by more than 13 dB relative to the GSC in reverberated environments. However, the AEC in GSAEC blocks the echo signal in the upper branch only and therefore the echo might leak to the beamformer output through the lower branch. The mild echo cancellation performance might be due to this leakage.

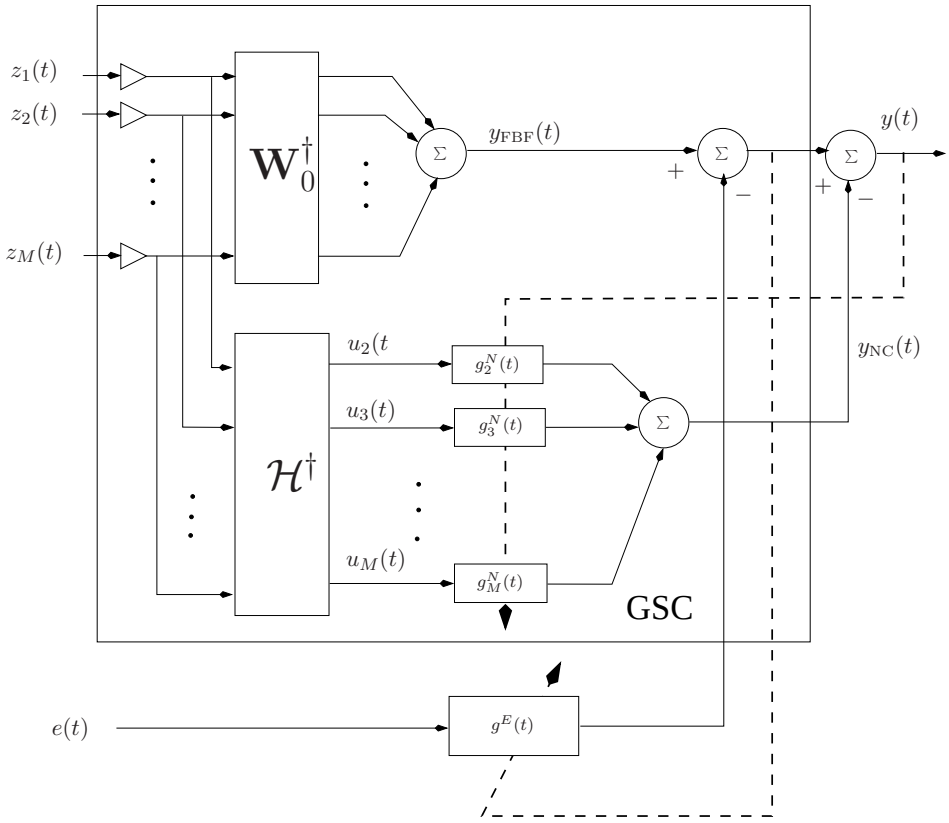


Figure 4.1: The GSAEC scheme

Affes and Grenier proposed in [13] a GSC structure for double talk situations. They presented a distortionless fixed beamformer constrained to cancel the echo, and a blocking matrix constrained to block both the desired signal and echo signal. The TFs are estimated using subspace tracking methods. These estimates are used to construct both the fixed beamformer and the blocking matrix. Nevertheless, subspace methods are known to be sensitive to the statistical distribution of the input signal. Furthermore, it is assumed in [13] that the noise signal is white, while in many cases this requirement cannot be met.

Two additional schemes for noise reduction and echo cancellation are proposed and compared in [15] by Doclo and Moonen. The first scheme includes a multi-channel AEC followed by a *generalized singular value decomposition* (GSVD) based beamformer. The second scheme incorporates the far-end echo reference directly into the GSVD beamformer, without cancelling the echo in every received signal. Simulations indicate that the first scheme outperforms the second one. Rombouts and Moonen [16] combine the speech enhancement and echo cancellation tasks in one integrated scheme. The optimization problem defined by this scheme is solved adaptively using a QRD-based least squares lattice (QRD-LSL) algorithm. It is shown that the performance of the integrated scheme is superior to the performance of traditional (cascading) schemes, while complexity is kept at an affordable level.

In this chapter, a modified TF-GSC scheme for joint noise reduction and echo cancellation is presented (denoted ETF-GSC), based on the structure proposed in [5]. The blocking matrix in the TF-GSC is modified to block the echo signal as well as the desired signal, while an AEC is incorporated to the adaptive noise canceller to reduce the echo from the MBF output. Compared to the DTF-GSC suggested in Chapter 2, better interference reduction can be obtained by incorporating the reference signal at hand. The proposed solution is evaluated through series of experiments and compared to two other schemes introduced in [12]. Experimental results demonstrate the performance of the proposed algorithms in noisy and reverberant environments in single and double talk situations.

The structure of this chapter is as follows. We formulate the problem of joint noise reduction and acoustic echo cancellation in Section 4.2. In Section 4.3 we repeat on main results obtained in [12]. A TF-GSC based scheme is suggested to jointly reduce noise and cancel echo in Section 4.4. In Section 4.5, we describe the experimental study and discuss the results. Finally, we conclude in Section 4.6.

## 4.2 Problem formulation

The received signal is comprised of three components, the desired signal source, echo signal and interference signal, as depicted in Figure 4.2.

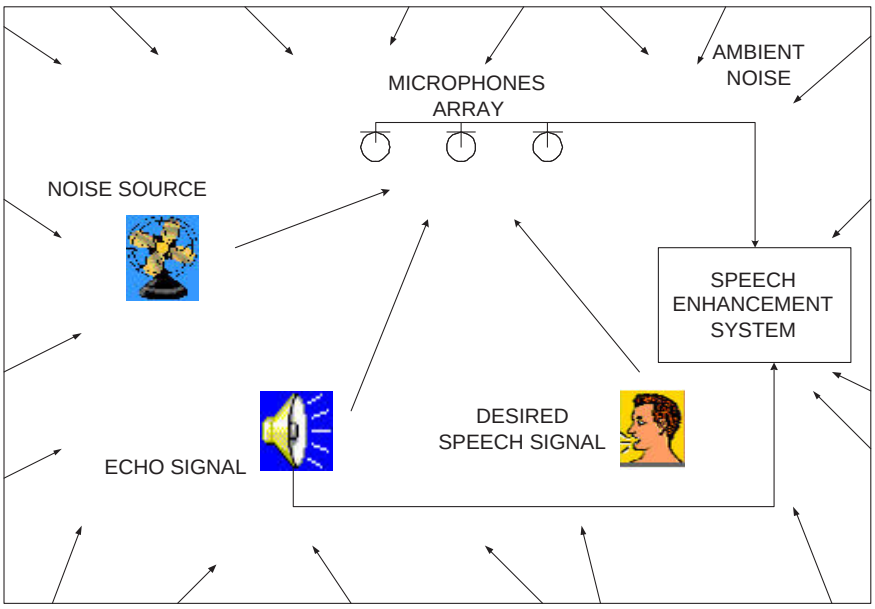


Figure 4.2: Desired and echo signals in noisy and reverberating environment

Assuming that the echo signal is measured at the loudspeaker, $s_2(t)$ in (2.1), the problem formulation for the general case, can be replaced with $e(t)$. Hence, the $m$-th microphone signal is

$$z_m(t) = a_m(t) * s(t) + b_m(t) * e(t) + n_m(t) \; ; \; m = 1, \ldots, M \tag{4.1}$$

where $a_m(t)$ is the impulse response of the filter relating the desired speech source and the $m$-th microphone; $s(t)$ is the desired signal source; $b_m(t)$ is the impulse response of the filter relating the echo speech source and the $m$-th microphone; $e(t)$ is the available echo signal measured at the loudspeaker and received by the microphone array through the acoustic path $b_m(t)$; $n_m(t)$ is the interference signal of the $m$-th microphone and $*$ denotes convolution. No separate measurement of the noise signal and the desired signal are available.

The analysis frame duration is chosen such that the signal may be considered stationary over the analysis frame. Typically, the impulse responses $a_m(t)$ and $b_m(t)$ are slowly changing in time and can be considered stationary over the analysis frame. In the *short time Fourier transform* (STFT) domain, (4.1) can be approximately rewritten as (for more details regarding the conditions for the approximation, please refer to footnote 2 in (2.2)):

$$\begin{aligned} Z_m(t,e^{j\omega}) \approx\ & A_m(e^{j\omega})S(t,e^{j\omega}) + B_m(e^{j\omega})E(t,e^{j\omega}) + N_m(t,e^{j\omega}) \\ & m = 1,\ldots,M \end{aligned} \tag{4.2}$$

where $Z_m(t,e^{j\omega}), S(t,e^{j\omega}), E(t,e^{j\omega})$ and $N_m(t,e^{j\omega})$ are the STFT of the respective signals. $A_m(e^{j\omega})$ and $B_m(e^{j\omega})$ are the ATFs from the desired source and echo source to the $m$-th microphone, respectively, which are assumed hereinafter time invariant over the observation period. A vector formulation of (4.2) is

$$\boldsymbol{Z}(t,e^{j\omega}) = \boldsymbol{A}(e^{j\omega})S(t,e^{j\omega}) + \boldsymbol{B}(e^{j\omega})E(t,e^{j\omega}) + \boldsymbol{N}(t,e^{j\omega}) \tag{4.3}$$

where

$$\begin{aligned} \boldsymbol{Z}(t,e^{j\omega}) &= \left[\, Z_1(t,e^{j\omega})\ Z_2(t,e^{j\omega})\ \cdots\ Z_M(t,e^{j\omega}) \,\right]^T \\ \boldsymbol{A}(e^{j\omega}) &= \left[\, A_1(e^{j\omega})\ A_2(e^{j\omega})\ \cdots\ A_M(e^{j\omega}) \,\right]^T \\ \boldsymbol{B}(e^{j\omega}) &= \left[\, B_1(e^{j\omega})\ B_2(e^{j\omega})\ \cdots\ B_M(e^{j\omega}) \,\right]^T \\ \boldsymbol{N}(t,e^{j\omega}) &= \left[\, N_1(t,e^{j\omega})\ N_2(t,e^{j\omega})\ \cdots\ N_M(t,e^{j\omega}) \,\right]^T . \end{aligned}$$

Our problem is to reconstruct the desired speech signal $S(t,e^{j\omega})$ (or a filtered version thereof) from the noisy observations $\boldsymbol{Z}(t,e^{j\omega})$. Hence, the noise should be reduced and the echo should be cancelled. However, opposed to the DTF-GSC suggested in Chapter 2, we wish to exploit the echo signal for better echo cancellation performance.

## 4.3 Cascade schemes

In this section we present two joint acoustic echo and noise cancellation schemes implemented in the frequency domain, based on handling each interference separately. Echo signals alone can be suppressed successfully by (AEC). In addition, background noise reduction can be obtained using multi-microphone signal enhancement techniques, e.g. fixed and adaptive beamforming algorithms. Hence, since it is required to reduce both acoustic echo and noise, solutions based on reducing each interference separately, are suggested. Similar to [30], we present two schemes each comprised of two components: beamformer and acoustic echo canceller. These schemes were presented in a shorter conference version [12]. Here we elaborate on the schemes and present a more comprehensive experimental study together with a comparison with the new proposed method. As depicted in Figures 4.3 and 4.4, the first scheme contains multi-channel AEC followed by a beamformer (denoted AEC-BF) and the second scheme contains beamformer followed by a single channel AEC (denoted BF-AEC). Note, that opposed to [30] which is implemented in the time domain, these blocks are implemented in the frequency domain. The frequency domain allows better convergence performance regardless of the condition number of the correlation matrix of the input data, and therefore is more suitable for speech processing. In both schemes, the frequency domain beamformer implemented using the TF-GSC [5], while the AEC is the BLMS algorithm.

### 4.3.1 AEC-BF scheme

When using the AEC-BF scheme, each AEC block receives an input signal which is comprised of the desired signal, echo signal and noise. Using the available remote speaker signal, $E(t, e^{j\omega})$, and an estimate of echo path $G_m^E(t, e^{j\omega})$, the AEC enhances the desired signal by cancelling the echo component. The following beamformer uses the enhanced signals to mitigate the noise by steering the array towards to the desired signal direction and removing the estimated noise

component, while assuming that the echo signal was already cancelled.

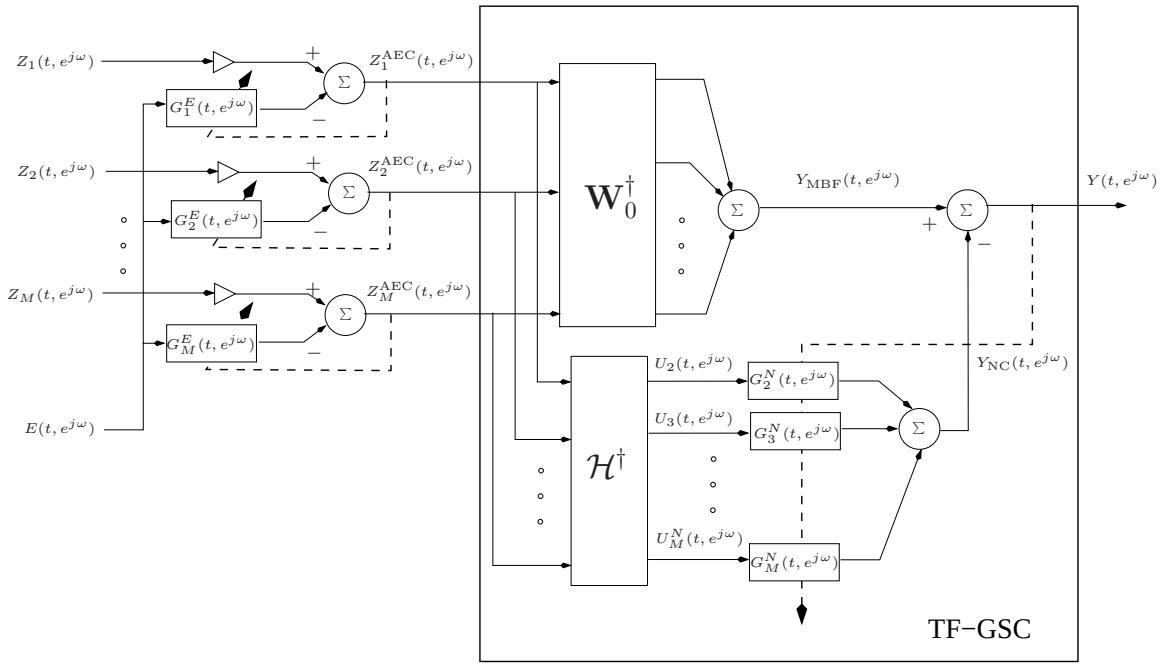


Figure 4.3: AEC-BF scheme

The multi-channel AEC output signals are

$$Z_m^{\mathrm{AEC}}(t,e^{j\omega}) = Z_m(t,e^{j\omega}) - G_m^{E\dagger}(t,e^{j\omega})E(t,e^{j\omega}), m = 1,\ldots,M; \tag{4.4}$$

where $G_m^E(t,e^{j\omega})$ is the $m$-th AEC filter. The filters are updated using the BLMS algorithm,

$$\begin{aligned} \tilde{G}_m^E(t+1,e^{j\omega}) &= G_m^E(t,e^{j\omega}) + \mu^E \frac{E(t,e^{j\omega})Z_m^*(t,e^{j\omega})}{P_{\mathrm{est}_m}^E(t,e^{j\omega})} \\ G_m^E(t+1,e^{j\omega}) &\xleftarrow{\mathrm{FIR}} \tilde{G}_m^E(t+1,e^{j\omega}) \end{aligned} \tag{4.5}$$

where $P_{\mathrm{est}_m}^E(t,e^{j\omega})$ is updated using,

$$P_{\mathrm{est}_m}^E(t,e^{j\omega}) = \eta^E P_{\mathrm{est}_m}^E(t-1,e^{j\omega}) + (1-\eta^E)|E(t,e^{j\omega})|^2 \tag{4.6}$$

while $\eta^E$ and $\mu^E$ are the step size and forgetting factor, respectively.

Nevertheless, when filtering is realized using multiplication in the frequency domain, aliasing effect due to cyclic convolution must be eliminated by imposing

an FIR constraint. Denoted as $\overset{\text{FIR}}{\longleftarrow}$, the FIR constraint includes the following three stages. First, we transform the multiplication result back to the time domain. Second, we truncate the resulting impulse response to the proper order. Third, we transform the resulting filter to the frequency domain.

We then apply the TF-GSC algorithm. The TF-GSC beamformer uses the multi-channel AEC output signals $\boldsymbol{Z}^{\text{AEC}}(t, e^{j\omega})$ as inputs. The matched beamformer output signal is

$$Y_{\text{MBF}}(t, e^{j\omega}) = \boldsymbol{W}_0^{\dagger}(e^{j\omega})\boldsymbol{Z}^{\text{AEC}}(t, e^{j\omega}) \tag{4.7}$$

where $\boldsymbol{W}_0(e^{j\omega})$ is the matched beamformer filter defined in [5]. $Y_{\text{NC}}(t, e^{j\omega})$, the estimated noise component in $Y_{\text{MBF}}$, is evaluated using

$$Y_{\text{NC}}(t, e^{j\omega}) = \boldsymbol{G}^{N\dagger}(t, e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{Z}^{\text{AEC}}(t, e^{j\omega}) \tag{4.8}$$

where the blocking matrix $\mathcal{H}(e^{j\omega})$ and adaptive noise canceller $\boldsymbol{G}^{N}(t, e^{j\omega})$ are defined in [5]. Therefore, the beamformer output signal is

$$Y(t, e^{j\omega}) = Y_{\text{MBF}}(t, e^{j\omega}) - Y_{\text{NC}}(t, e^{j\omega}). \tag{4.9}$$

Similarly to the echo cancellation, only non-aliased samples are kept while calculating the output signal of the beamformer.

It is well known that the AEC performance is sensitive to noise presence. When the multi-channel AEC precede the beamformer, the AEC input signals are contaminated by noise and therefore the ability to reduce echo degrades. On the other hand, the input signals of the BF are free of echo in this scheme, and therefore the directivity of the ANC is used to cancel the noise only.

### 4.3.2 BF-AEC scheme

When using the BF-AEC scheme, the BF receives input signals which are comprised of the desired signal, echo signal and noise. The beamformer enhances the desired signal by reducing the noise from the input signals. Using the available remote speaker signal $E(t, e^{j\omega})$ and an estimate of echo path $G^{E}(t, e^{j\omega})$, the following AEC enhances the desired signal by cancelling the echo component.

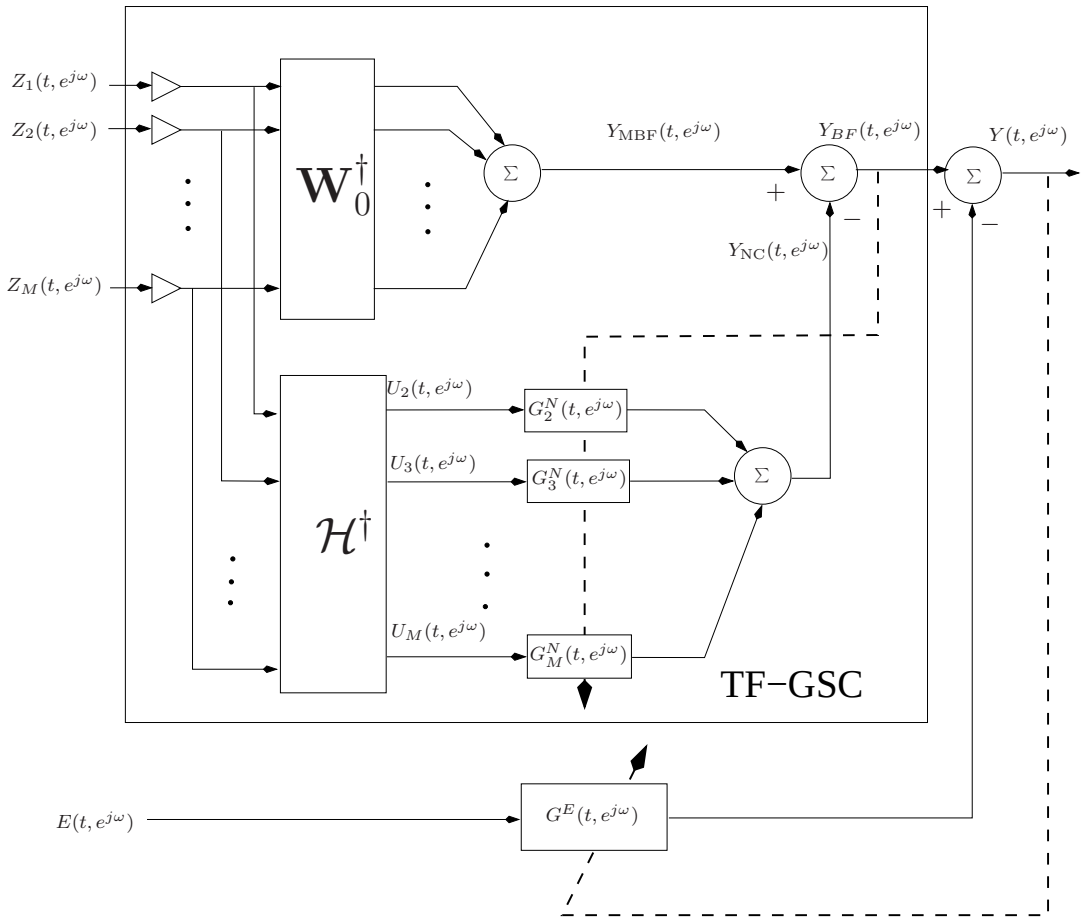


Figure 4.4: BF-AEC scheme

We first apply the TF-GSC algorithm. The matched beamformer output signal is

$$Y_{\mathrm{MBF}}(t, e^{j\omega}) = \boldsymbol{W}_0^{\dagger}(e^{j\omega})\boldsymbol{Z}(t, e^{j\omega}). \tag{4.10}$$

$Y_{\mathrm{NC}}(t, e^{j\omega})$, the estimated noise component in $Y_{\mathrm{MBF}}$, is evaluated using

$$Y_{\mathrm{NC}}(t, e^{j\omega}) = \boldsymbol{G}^{N\dagger}(t, e^{j\omega})\mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t, e^{j\omega}) \tag{4.11}$$

and the beamformer output signal is

$$Y_{\mathrm{BF}}(t, e^{j\omega}) = Y_{\mathrm{MBF}}(t, e^{j\omega}) - Y_{\mathrm{NC}}(t, e^{j\omega}). \tag{4.12}$$

The following single channel AEC uses the beamformer output as input signal. The AEC output signal is

$$Y(t, e^{j\omega}) = Y_{\mathrm{BF}}(t, e^{j\omega}) - G^{*E}(t, e^{j\omega})E(t, e^{j\omega}), \tag{4.13}$$

where $G^E(t, e^{j\omega})$ is the AEC filter. The filters are updated using the BLMS algorithm,

$$\begin{aligned} \tilde{G}^E(t+1, e^{j\omega}) &= G^E(t, e^{j\omega}) + \mu^E \frac{E(t, e^{j\omega})Y^*(t, e^{j\omega})}{P^E_{\mathrm{est}}(t, e^{j\omega})} \\ G^E(t+1, e^{j\omega}) &\xleftarrow{\mathrm{FIR}} \tilde{G}^E(t+1, e^{j\omega}) \end{aligned} \tag{4.14}$$

where $\xleftarrow{\mathrm{FIR}}$ denotes the FIR constrain defined in Section 4.3.1 and $P^E_{\mathrm{est}}(t, e^{j\omega})$ is updated using

$$P^E_{\mathrm{est}}(t, e^{j\omega}) = \eta^E P^E_{\mathrm{est}}(t-1, e^{j\omega}) + (1-\eta^E)|E(t, e^{j\omega})|^2. \tag{4.15}$$

Similarly to the previous scheme, only non-aliased samples are kept while calculating the output signal of the beamformer and the AEC.

When the beamformer precede the AEC, the echo cancellation performance is not impaired by the noise, since the beamformer reduced it. However, the performance of the AEC may deteriorate due to the time variations in the echo path caused by the beamformer. Note that when both noise and echo are present, the TF-GSC can eliminate both due to its directivity.

## 4.4 ETF-GSC scheme

As presented in Section 4.3, when using the AEC-BF scheme, the AEC performance impaired significantly due to the noise at the received signals. Furthermore, the performance of the AEC in the BF-AEC scheme severely deteriorates due to the time variations of the echo path caused by the beamformer. Hence, both schemes may suffer from disadvantages, and a combined solution for reducing both noise and echo is required.

In this section we will show that noise and echo can be reduced jointly by some modifications to the TF-GSC scheme. The suggested scheme contains three blocks: matched beamformer denoted here $\boldsymbol{F}_0^{\dagger}(t, e^{j\omega})$, blocking unit (BU) and a multi-channel adaptive noise and echo canceller (ANEC). The MBF is designed to maintain the desired signal direction, as the MBF in the TF-GSC (no blocking of the echo). The BU blocks both desired signal and echo signal directions, by using two sub-blocks. The first, $\mathcal{H}^{\dagger}(e^{j\omega})$ blocks the desired signal direction, similarly to the BM in the TF-GSC, while the second, multi-channel AEC filters $\boldsymbol{G}^{H}(t, e^{j\omega})$ cancel the echo in $\boldsymbol{U}(t, e^{j\omega})$ in order to create the reference signals $\boldsymbol{U}'(t, e^{j\omega})$. Updating $\boldsymbol{G}^{H}(t, e^{j\omega})$ is done using the beamformer output $Y(t, e^{j\omega})$, since it contains less noise than $\boldsymbol{U}'(t, e^{j\omega})$. The ANEC uses the resulting reference signals $\boldsymbol{U}'(t, e^{j\omega})$ and $E(t, e^{j\omega})$ to reduce the noise and suppress the echo, respectively. The ETF-GSC scheme is depicted in Figure 4.5.

The matched beamformer maintains the desired direction and therefore its output signal $Y_{\mathrm{MBF}}(t, e^{j\omega})$ contains steered desired signal component, echo and noise components. The treatment of the two interference signals is decoupled into two sub-blocks. By blocking both desired and echo signals, the BU produces the reference signals $\boldsymbol{U}'(t, e^{j\omega})$. The ANEC uses adaptive filters $\boldsymbol{G}^{N}(t, e^{j\omega})$ and $\boldsymbol{U}'(t, e^{j\omega})$ to reduce the noise component in $Y_{\mathrm{MBF}}(t, e^{j\omega})$, as in the TF-GSC. The lower sub-block $\boldsymbol{G}^{E}(t, e^{j\omega})$ uses the echo reference signal $E(t, e^{j\omega})$ to cancel echo.

Note that filtered versions of $E(t, e^{j\omega})$ are used as reference signals to $\boldsymbol{G}^{E}(t, e^{j\omega})$ and $\boldsymbol{G}^{H}(t, e^{j\omega})$. We will elaborate on this usage now. Updating $\boldsymbol{G}^{E}(t, e^{j\omega})$ is done

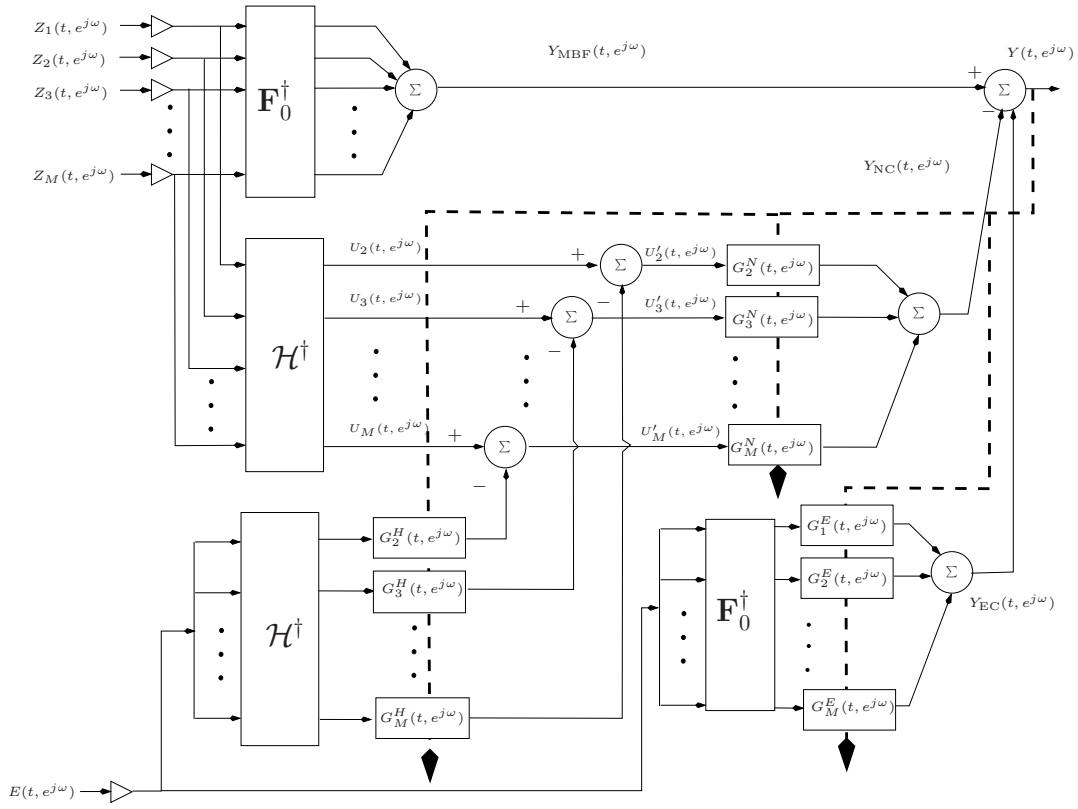


Figure 4.5: ETF-GSC scheme

using the beamformer output $Y(t, e^{j\omega})$, while the echo component in the beamformer output is the sum

$$\sum_{m=1}^{M} E(t, e^{j\omega}) F_{0_m}^{*}(t, e^{j\omega}) B_m(e^{j\omega}) = E(t, e^{j\omega}) \boldsymbol{F}_0^{\dagger}(t, e^{j\omega}) \boldsymbol{B}(e^{j\omega}) \tag{4.16}$$

where $F_{0_m}(t, e^{j\omega})$ are the components of $\boldsymbol{F}_0(t, e^{j\omega})$. Therefore, the TF which relates the remote speaker source $E(t, e^{j\omega})$ and the beamformer output is $\boldsymbol{F}_0^{\dagger}(t, e^{j\omega})\boldsymbol{B}(e^{j\omega})$. In other words, if a single channel AEC is used to suppress the echo, feeded by $E(t, e^{j\omega})$ as reference signal, the filter will converge to $\boldsymbol{F}_0^{\dagger}(t, e^{j\omega})\boldsymbol{B}(e^{j\omega})$. Note that this TF is time variant and depends on known MBF filters $\boldsymbol{F}_0(t, e^{j\omega})$. However, it is essential to keep the AEC filters as short as possible for fast convergence, and avoid unnecessary tracking. Therefore, by using $E(t, e^{j\omega})\boldsymbol{F}_{0_m}^{*}(t, e^{j\omega})$, $m = 1, \ldots, M$; as reference signals for a multi channel AEC, the $m$-th filter converges to $B_m(e^{j\omega})$ alone. Hence, the echo reference signal is replicated $M$ times and filtered by the available MBF filters to create the necessary reference signals to $\boldsymbol{G}^E(t, e^{j\omega})$. For similar reasons, filtered versions of $E(t, e^{j\omega})$ are used as reference signals to $\boldsymbol{G}^H(t, e^{j\omega})$.

Note that the filter sets $\boldsymbol{G}^H(t, e^{j\omega})$, $\boldsymbol{G}^N(t, e^{j\omega})$ and $\boldsymbol{G}^E(t, e^{j\omega})$ operate separately. It might have different length and different step size, according to the problem at hand. Moreover, $\boldsymbol{G}^N(t, e^{j\omega})$ adapts during noise only frames, while $\boldsymbol{G}^H(t, e^{j\omega})$ and $\boldsymbol{G}^E(t, e^{j\omega})$ adapt during echo frames only. Detecting activity in the echo signal is done using the reference signal $e(t)$, whereas perfect VAD is assumed for the desired signal.

The estimation of the blocks is done in a similar way to the estimation previously described in Section 2.5. As in the TF-GSC, the MBF and BM require estimation of $\frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}$, which can be done during desired signal frames. The algorithm is summarized in Figure 4.6.

1) Matched beamformer:

$$Y_{\mathrm{MBF}}(t, e^{j\omega}) = \boldsymbol{F}_0^{\dagger}(e^{j\omega})\boldsymbol{Z}(t, e^{j\omega})$$

where $\boldsymbol{F}_0(t, e^{j\omega}) = \frac{\boldsymbol{H}(e^{j\omega})}{\|\boldsymbol{H}(e^{j\omega})\|^2}\mathcal{F}(e^{j\omega})$ and $\boldsymbol{H}(e^{j\omega}) = \frac{\boldsymbol{A}(e^{j\omega})}{A_1(e^{j\omega})}$.

2) Noise reference signals:

$$\boldsymbol{U}(t, e^{j\omega}) = \mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{Z}(t, e^{j\omega})$$
$$\boldsymbol{U}'(t, e^{j\omega}) = \boldsymbol{U}(t, e^{j\omega}) - E(t, e^{j\omega})\boldsymbol{P}(t, e^{j\omega})$$

3) NR output signal:

$$Y_{\mathrm{NC}}(t, e^{j\omega}) = G^{N\dagger}(t, e^{j\omega})\boldsymbol{U}(t, e^{j\omega})$$

4) AEC output signal:

$$Y_{\mathrm{EC}}(t, e^{j\omega}) = G^{E\dagger}(t, e^{j\omega})E(t, e^{j\omega})F_0^*(e^{j\omega})$$

5) Output signal:

$$Y(t, e^{j\omega}) = Y_{\mathrm{MBF}}(t, e^{j\omega}) - (Y_{\mathrm{NC}}(t, e^{j\omega}) + Y_{\mathrm{EC}}(t, e^{j\omega}))$$

6) BU Filters update, for $m = 2, \ldots, M$:

$$\tilde{G}_m^H(t+1, e^{j\omega}) = G_m^H(t, e^{j\omega}) + \mu^H \frac{U_m(t, e^{j\omega})Y^*(t, e^{j\omega})}{P_{\mathrm{est}}^H(t, e^{j\omega})}$$
$$G_m^H(t+1, e^{j\omega}) \xleftarrow{\mathrm{FIR}} \tilde{G}_m^H(t+1, e^{j\omega})$$

where, $P_{\mathrm{est}}^H(t, e^{j\omega}) = \eta^H P_{\mathrm{est}}^H(t-1, e^{j\omega}) + (1-\eta^H)\sum_m |(E(t, e^{j\omega})\mathcal{H}(e^{j\omega}))_m|^2$

7) NR Filters update, for $m = 2, \ldots, M$:

$$\tilde{G}_m^N(t+1, e^{j\omega}) = G_m^N(t, e^{j\omega}) + \mu^N \frac{U_m(t, e^{j\omega})Y^*(t, e^{j\omega})}{P_{\mathrm{est}}^N(t, e^{j\omega})}$$
$$G_m^N(t+1, e^{j\omega}) \xleftarrow{\mathrm{FIR}} \tilde{G}_m^N(t+1, e^{j\omega})$$

where, $P_{\mathrm{est}}^N(t, e^{j\omega}) = \eta^N P_{\mathrm{est}}^N(t-1, e^{j\omega}) + (1-\eta^N)\sum_m |U'_m(t, e^{j\omega})|^2$

8) EC Filters update, for $m = 1, \ldots, M$:

$$\tilde{G}_m^E(t+1, e^{j\omega}) = G_m^E(t, e^{j\omega}) + \mu^E \frac{F_{0_m}^*(e^{j\omega})E(t, e^{j\omega})Y^*(t, e^{j\omega})}{P_{\mathrm{est}}^E(t, e^{j\omega})}$$
$$G_m^E(t+1, e^{j\omega}) \xleftarrow{\mathrm{FIR}} \tilde{G}_m^E(t+1, e^{j\omega})$$

where, $P_{\mathrm{est}}^E(t, e^{j\omega}) = \eta^E P_{\mathrm{est}}^E(t-1, e^{j\omega}) + (1-\eta^E)\sum_m |F_{0_m}^*(e^{j\omega})E(t, e^{j\omega})|^2$

9) keep only non-aliased samples.

(note: $\mathcal{H}(e^{j\omega})$ is defined in (2.32)).

Figure 4.6: Summary of the ETF-GSC.

## 4.5 Experimental study

### 4.5.1 Setup

The proposed algorithms were tested in a simulated room environment. The desired and echo speech signals were drawn from the TIMIT [24] database, while a speech-like noise from NOISEX-92 [25] database was used to simulate directional stationary noise source. All three signals were filtered by simulated room impulse responses, resulting in directional signals, which are received by $M = 10$ microphones. Allen and Berkley's *image method* [26] was used to simulate the ATFs with reverberation time set to $T_{60} = 40$ms (see Fig. 4.7 for a typical impulse and frequency responses of the acoustical path). The sampling frequency was 8KHz and the resolution was set to 16 bits per sample.

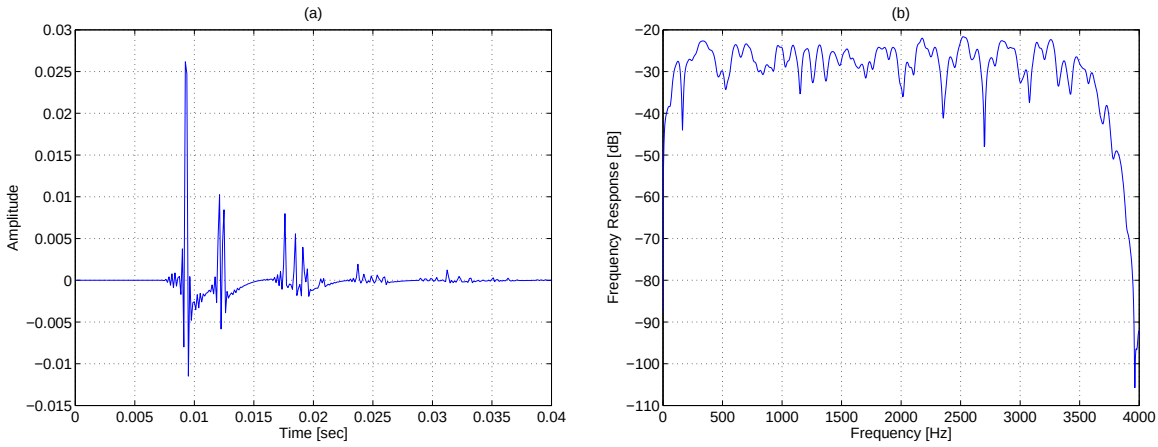


Figure 4.7: (a) Impulse response and (b) frequency response of the ATF between the desired source and the first microphone.

For the cascade schemes we use non casual FIR models in all filters. In the AEC, 500 coefficients were used, in the blocking filters of the TF-GSC the total filter length was set to 181 and filters of the interference cancellers to 251. Segments of 2048 samples were used to implement the *overlap and save* procedure.

For the ETF-GSC scheme, the lengths of the filters in the MBF, blocking matrix, echo cancellers and the noise cancellers were set to 150, 250, 400 and 350

taps, respectively. Two-sided FIR models were used for all filters. Segments of 2048 samples were used to implement the *overlap and save* procedure.

### 4.5.2 Results

The results for the AEC-BF scheme and BF-AEC scheme, in the presence of directional noise in various cases of input $SNR$ and $SER$ are depicted in Table 4.1 and 4.2, respectively. Two figures of merit are measured in different stages to evaluate the performance of the proposed schemes, the desired signal to noise ratio $SNR$ and desired signal to echo ratio $SER$. In cascade schemes, the SNR and SER are measured at the first microphone signal, at the output of the first stage and at the output of the second stage (total output). The measurements are taken in a time frame consisting of both echo and desired signal. The improvement in $SNR$ and $SER$, denoted as Noise Reduction and Echo suppression, respectively.

| Input | | Echo suppression | | | Noise reduction |
|---|---|---|---|---|---|
| SNR | SER | AEC | BF | Total | Total |
| 5 | 5 | 13.2 | 0.2 | 13.4 | 23.2 |
| 10 | 5 | 16.6 | 1.5 | 18.1 | 23.5 |
| 15 | 5 | 18.5 | 2.1 | 20.6 | 23.5 |
| 5 | 10 | 8.8 | 0.3 | 9.1 | 23.3 |
| 10 | 10 | 12.9 | 1.3 | 14.3 | 23.7 |
| 15 | 10 | 15.9 | 1.5 | 17.4 | 23.8 |
| 5 | 15 | 3.9 | 0.4 | 4.4 | 23.3 |
| 10 | 15 | 8.5 | 1.2 | 9.7 | 23.8 |
| 15 | 15 | 12.1 | 1.0 | 13.1 | 23.9 |

Table 4.1: AEC-BF scheme performance

Figure 4.8 shows waveforms for the ETF-GSC algorithm in the presence of directional noise field. In (a) and (b) a segment of the desired and echo signals are illustrated, respectively. Double talk situation is clearly observed. The signal measured at microphone # 1, which comprises the desired speech, the echo and a directional noise signal, is depicted in (c). The respective enhanced signal, after the algorithm has adapted, is depicted in (d). The experiment was conducted while input $SNR = 5dB$ and $SER = 5dB$.

| Input | | Echo suppression | | | Noise reduction |
|---|---|---|---|---|---|
| SNR | SER | AEC | BF | Total | Total |
| 5 | 5 | 2.6 | 7.4 | 9.9 | 23.7 |
| 10 | 5 | 4.7 | 6.6 | 11.3 | 24.7 |
| 15 | 5 | 5.4 | 6.2 | 11.7 | 24.7 |
| 5 | 10 | 1.6 | 7.4 | 8.9 | 24.1 |
| 10 | 10 | 3.7 | 6.9 | 10.6 | 24.5 |
| 15 | 10 | 4.5 | 6.6 | 11.1 | 24.5 |
| 5 | 15 | 1.7 | 6.8 | 8.6 | 24.4 |
| 10 | 15 | 3.8 | 6.2 | 10.0 | 24.6 |
| 15 | 15 | 4.6 | 5.8 | 10.4 | 24.5 |

Table 4.2: BF-AEC scheme performance

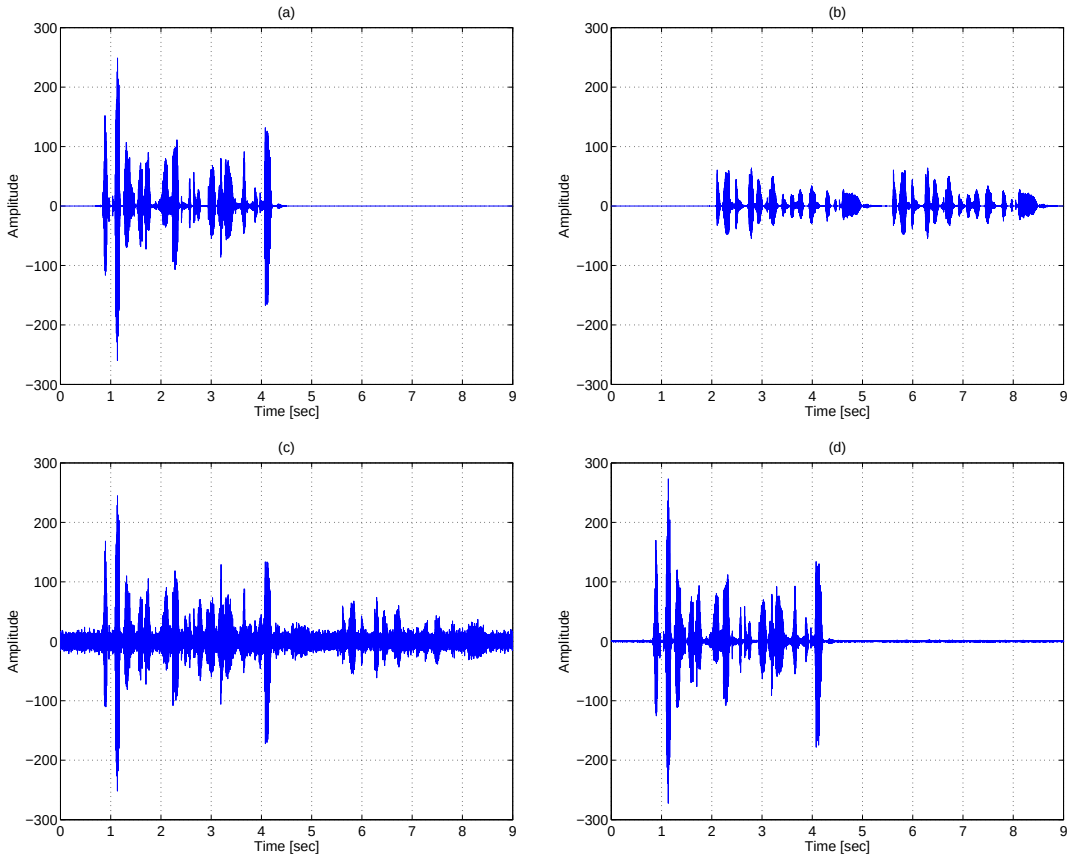


Figure 4.8: Speech waveforms ($SNR = 5dB$, $SER = 5dB$): (a) Desired signal, (b) Echo signal, (c) Mic. #1 signal, (d) ETF-GSC enhanced signal.

In the ETF-GSC scheme, SNR and SER are measured at the first microphone signal and at the output of the system. The results for directional noise are depicted in Table 4.3 for various cases of input $SNR$ and $SER$. The figures of merit were calculated for the time segment depicted in Figures 4.8.

| Input | | Echo suppression | Noise reduction |
|---|---|---|---|
| SNR | SER | Total | Total |
| 5 | 5 | 29.1 | 20.4 |
| 10 | 5 | 35.8 | 22.4 |
| 15 | 5 | 37.7 | 23.1 |
| 5 | 10 | 26.0 | 22.8 |
| 10 | 10 | 33.6 | 22.5 |
| 15 | 10 | 35.6 | 22.3 |
| 5 | 15 | 18.1 | 23.0 |
| 10 | 15 | 26.4 | 23.5 |
| 15 | 15 | 29.7 | 23.6 |

Table 4.3: ETF-GSC scheme performance

Figure 4.9 shows sonograms of the data depicted in Figure 4.8. It can be seen that for the directional noise field, both noise and interference signals are well suppressed, especially in frequencies above 500 Hz. Moreover, no self cancellation or other deviation can be noticed, even during the double talk situation.

### 4.5.3 Discussion

Analysis of the obtained results shows that the AEC-BF scheme almost always outperforms the BF-AEC scheme when comparing the echo suppression, while the ETF-GSC scheme outperform both. When considering noise reduction performance, all three schemes achieve similar results, although the BF-AEC slightly outperforms the others. This phenomenons need further discussion. We will first discuss the echo cancellation performance and then the noise reduction.

When noise and echo signals are present, the TF-GSC can eliminate both due to its directivity. However, the performance of the following AEC severely deteriorates due to the time variations of the echo path caused by the beamformer. Note the number of taps in the AEC filters, chosen according to the

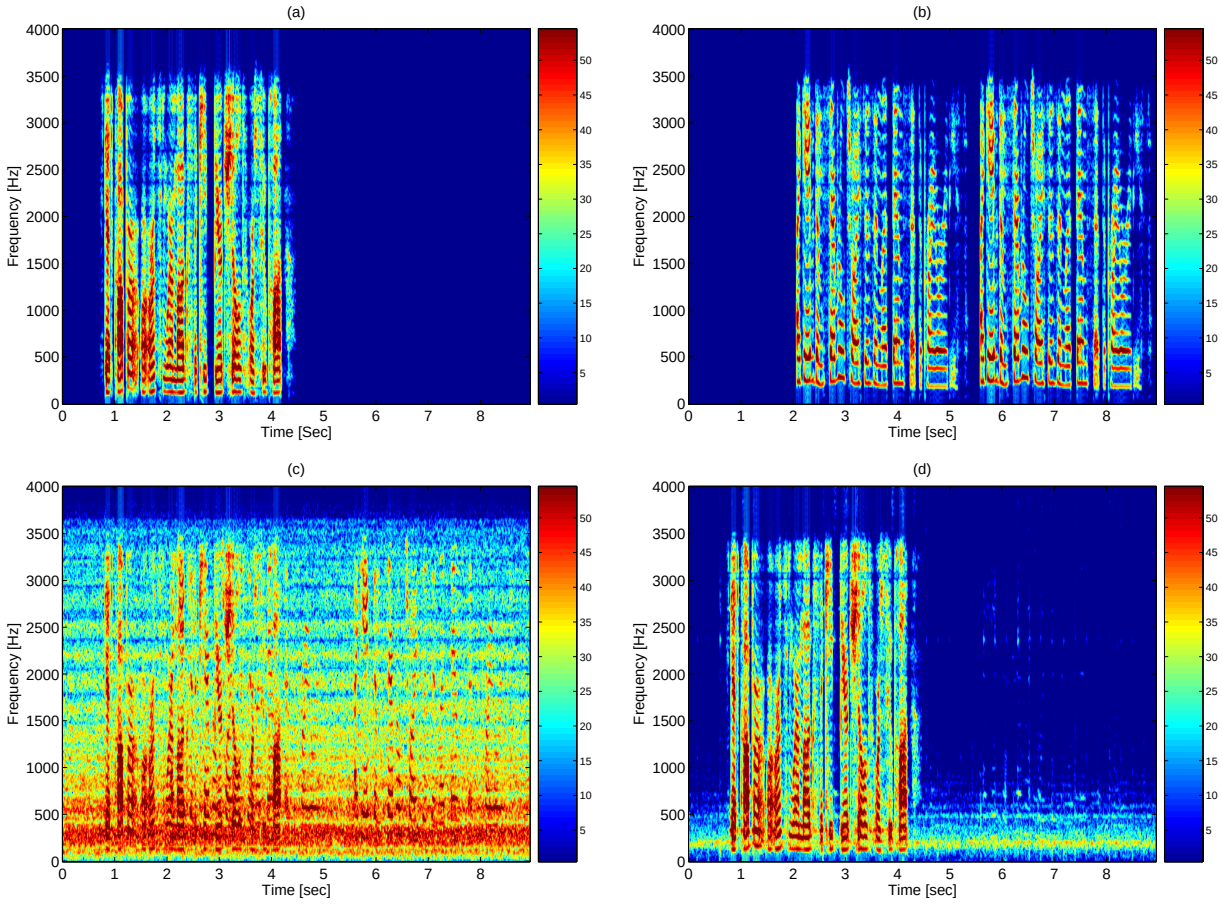


Figure 4.9: Sonograms ($SNR = 5dB$, $SER = 5dB$): (a) Desired signal, (b) Echo signal, (c) Mic. #1 signal, (d) ETF-GSC enhanced signal.

acoustical environment, is usually greater than the number of taps required for the ANC. Therefore, the AEC converges slower and its ability to track changes is limited. On the other hand, while the AEC precede the beamformer, the degradation of the AEC performance due to the existence of noise signals is partially compensated by the following beamformer. For that reason the AEC-BF scheme almost always outperforms the BF-AEC scheme when comparing the total echo suppression performance.

The echo suppression performance of the ETF-GSC scheme outperforms considerably both the AEC-BF scheme and BF-AEC scheme. For example, the ETF-GSC scheme achieves $37.7dB$ echo suppression when $SNR = 15dB$ and $SER = 5dB$, while in the same environment conditions, the AEC-BF and the BF-AEC suppress the echo by $20.6dB$ and $11.7dB$, respectively. The ETF-GSC enjoys the benefits of the two cascade schemes. First, the convergence of the AEC filters in the ETF-GSC scheme is not impaired due to noise presence, since the error feedback is taken from the output signal, after the noise was reduced. Second, since the AEC in the ETF-GSC filters need to estimate the ATF $\boldsymbol{B}(e^{j\omega})$ only (see Section 4.4), the convergence is faster compared to BF-AEC scheme.

Note that when the input signals contain dominant presence of echo and negligible noise, the BF in the BF-AEC achieves considerable echo suppression. For example, when the $SNR = 15dB$ and $SER = 5dB$ the BF achieves 5.4 dB echo suppression, while only 1.7 dB echo suppression is obtained when $SNR = 5dB$ and $SER = 15dB$. This is due to the significant component of echo in the reference signals produced by the blocking matrix. This demonstrates the ability of the BF to suppress echo using its directivity. Nevertheless, when the BF directivity is exploited to cancel echo, two nulls must be implemented by the ANC, first pointing to the noise direction and second to echo direction. Therefore the noise reduction performance deteriorates.

It should be noted that AEC-BF echo performance significantly deteriorates in the presence of dominant noise and negligible echo signals. For example, the echo suppression while $SNR = 5dB$ and $SER = 15dB$ is 4.4dB only. Since the echo

components are masked by noise, the AEC filters are not able to use the output signal for error correction and tracking. On the other hand, in the BF-AEC scheme the single channel AEC uses enhanced signal and achieves 8.6dB echo suppression and the joint system in the ETF-GSC reduces the echo by 18.1dB.

When comparing the noise reduction performance, all three schemes demonstrate similar performance, an average of 23dB noise attenuation for all tested cases. Nevertheless, the BF-AEC slightly outperforms the AEC-BF and the ETF-GSC schemes. While the noise reduction in both AEC-BF and ETF-GSC is 23dB in average, the BF-AEC achieves approximately 24dB noise attenuation. Although the BF in the AEC-BF uses echo suppressed signals and therefore the ANC is utilized toward noise reduction only, it achieves inferior noise reduction performance compared to the BF-AEC. This is due to the time variations imposed by the AEC on the total TF which relates the directional noise source and the BF inputs.

Note that the ETF-GSC results were obtained in a double talk situation, while the cascade schemes were tested in single talk situation only, due to its significant performance degradation in the presence of double talk. As an example, we will now explain the reasons for the performance degradation in the BF-AEC scheme. Let us consider the AEC filter $G^E(t, e^{j\omega})$, as depicted in Figure 4.4. Note that the output of the scheme, $Y(t, e^{j\omega})$, is used as error correction feedback for updating the AEC filter $G^E(t, e^{j\omega})$. Hence, when the AEC adapts during double talk frames, the desired signal component in $Y(t, e^{j\omega})$ deteriorates the AEC convergence.

Furthermore, when using the TF-GSC beamformer in the presence of nonstationary interference, such as echo signal, it can be shown that the noise reduction performance impaired considerably. Using the problem formulated in (4.3) we obtain

$$\begin{aligned} Z_m(t, e^{j\omega}) &= A_m(e^{j\omega})S(t, e^{j\omega}) + B_m(e^{j\omega})E(t, e^{j\omega}) + N_m(t, e^{j\omega}) \\ & m = 1, \ldots, M. \end{aligned} \tag{4.17}$$

By substituting (4.17) in (20) [5], the obtained $m$-th reference signal is

$$U_m(t,e^{j\omega}) = Z_m(t,e^{j\omega}) - \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}Z_1(t,e^{j\omega}) \tag{4.18}$$
$$= E(t,e^{j\omega})\left[B_m(e^{j\omega}) - \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}B_1(e^{j\omega})\right] + N_m(t,e^{j\omega}) - \frac{A_m(e^{j\omega})}{A_1(e^{j\omega})}N_1(t,e^{j\omega}),$$

and the first received signal is

$$Z_1(t,e^{j\omega}) = A_1(e^{j\omega})S(t,e^{j\omega}) + B_1(e^{j\omega})E(t,e^{j\omega}) + N_1(t,e^{j\omega}). \tag{4.19}$$

Now, since $E(t,e^{j\omega})$ is nonstationary signal over the analysis frame, $\phi^{(k)}_{u_m z_1}(e^{j\omega})$, the cross-PSD between $U_m(t,e^{j\omega})$ and $Z_1(t,e^{j\omega})$ cannot be assumed independent with the index k. Therefore, the equation set presented in [5] (30), based on

$$\hat{\Phi}^{(k)}_{z_m z_1}(e^{j\omega}) = H_m(e^{j\omega})\hat{\Phi}^{(k)}_{z_1 z_1}(e^{j\omega}) + \Phi_{u_m z_1}(e^{j\omega}) + \varepsilon^{(k)}_m(e^{j\omega}),$$

is no longer correct. In that case, the ATFs' ratio estimation procedure impaired considerably and the fixed beamformer and blocking matrix as well.

However, when using the proposed ETF-GSC, the AEC error feedback is taken from the output signal, after the noise was reduced. Therefore, the convergence of the AEC filters in the ETF-GSC scheme is not impaired due to noise presence. Furthermore, since the echo was removed from the reference signals to the ANC, the equation set introduced in [5] (30) remains correct, hence proper ATFs' ratio estimation can be obtained.

## 4.6 Conclusion

In this chapter, we addressed the problem of joint echo cancellation and noise reduction in a reverberated environment. The competing speech signal in the problem formulated in Chapter 2 for the general case, is replaced with an echo signal. A solution based on the TF-GSC is presented. This can be done, by modifying some of the blocks and incorporating the reference signal to the echo as another input to the system. This scheme is evaluated through a series of

experiments in single and double talk situations, and compared to two additional schemes, implemented in the frequency domain.

Opposed to the cascade schemes, it is demonstrated that no impairment in the echo cancellation performance due to the noise presence or as a result of the beamforming process can be observed in the ETF-GSC scheme. Therefore, TF-GSC based scheme outperforms significantly the two other schemes while comparing the echo cancellation performance. Similar noise reduction performance are obtained in all three schemes.

# Chapter 5

# Conclusion

## 5.1 Summary

We presented a dual source interference canceller, based on the TF-GSC, for removing non-stationary directional interference and stationary interferences. The new structure was designated dual transfer function GSC (DTFGSC). As the regular GSC beamformer it is comprised of 3 blocks: matched beamformer (MBF), blocking matrix BM and adaptive noise canceller (ANC). The MBF and the blocking matrix were modified to handle the dual source case. A new system identification procedure was derived for estimating the blocking matrix terms directly during double talk segments. The proposed system may be applied to many interesting problems. One possible application is the BSS problem of convolutive mixtures and additive noise. The two sources can be extracted by exchanging the roles of the desired and competing speech signals.

As a result of signal source movement, change in the acoustic environment or estimation errors, the DTF-GSC algorithm might use inaccurate ATFs. We evaluated the degradation in performance due to these estimation errors. We have derived a general expression for the output power spectral density. From this expression we derived expressions for the expected deviation imposed on the desired signal and the amount of achievable noise and interference reduction. Gannot *et al.* [28] presented an analysis of the PSD deviation imposed on the desired signal by the TF-GSC algorithm. However, the structure of the matched

beamformer and blocking matrix in the DTF-GSC differs from the TF-GSC since both are constrained to block the competing speech signal direction. In addition, we present a noise reduction and interference reduction analysis as well.

Finally, we addressed the problem of joint echo cancellation and noise reduction in a reverberated environment. The competing speech signal in the problem formulated for the general case, is replaced with an echo signal. A solution based on the TF-GSC is presented (denoted ETF-GSC). This can be done, by modifying some of the blocks and incorporating the reference signal to the echo as another input. This scheme was evaluated through a series of experiments in single and double talk situations, and compared to two additional cascade schemes, implemented in the frequency domain. It was shown that the DTF-GSC scheme significantly outperforms the two others in echo cancellation performance. Similar noise reduction performance are obtained in all three schemes.

## 5.2 Future Research

The method we proposed in this work opens a number of interesting topics for future study:

### 5.2.1 Dual nonstationary speech signals in the presence of echo and stationary noise

In Chapter 2 we assumed an environment where the desired signal is contaminated by both competing nonstationary speech signal and stationary noise signal, while no reference signal is available. On the other hand, in Chapter 4 we assumed the received signal is comprised of three components, the desired signal, echo signal and stationary noise signal. However, in some cases, the acoustic environment comprises of two speech sources, echo source and stationary noise source. For example, in hands-free system located in a car, the received signal might be contaminated by competing speech originating from the back seat, echo signal from the loudspeaker and some stationary noise. The problem formulation is

depicted in Figure 5.1.

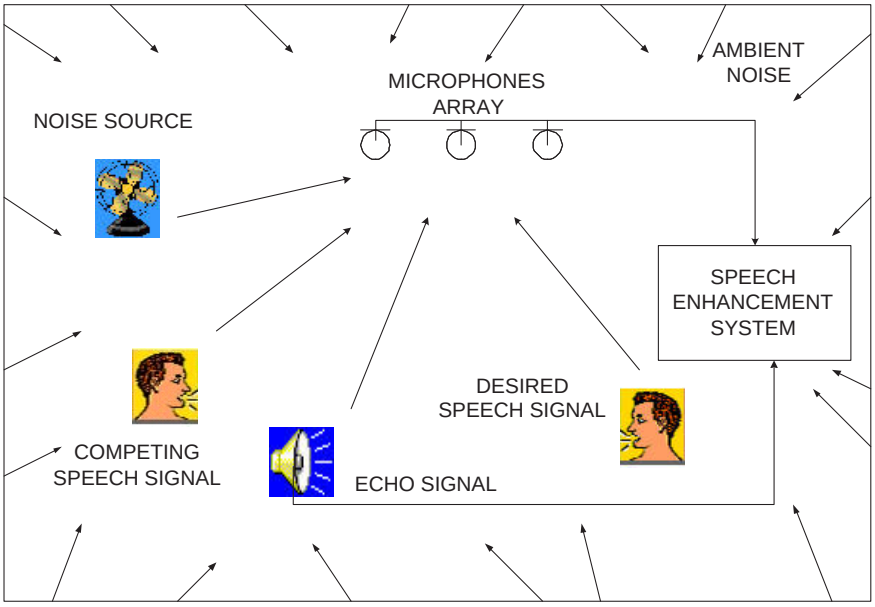


Figure 5.1: Dual nonstationary speech signals in the presence of echo and stationary noise

In this case, the DTF-GSC suggested in Chapter 2 can be extended to cancel the echo signal as well, by exploiting the reference signal at hand. Future research may include evaluation of the degradation in performance due to estimation errors.

### 5.2.2 Speech enhancement using the Dual Transfer Function GSC (DTF-GSC) and postfiltering

In speech enhancement applications microphone array postfiltering allows additional reduction of noise components at a beamformer output. The dual transfer function GSC (TF-GSC), proposed in Chapter 2 demonstrates impressive noise and nonstationary interference reduction performance in the presence of directional noise field, while maintaining low speech distortion. However, in a diffused noise field less significant noise reduction is obtained. Future research may pro-

pose postfiltering methods for improving the nonstationary interference reduction performance of the DTF-GSC. This can be done by making use of known algorithms concatenated to the DTF-GSC beamformer, or by exploiting the noise-only components constructed within the DTF-GSC structure. The improvement in the beamformer performance, should be analyzed and examined in various noise field types using objective and subjective quality measures.

### 5.2.3 Dual Transfer Function GSC (DTF-GSC) using Relative Transfer Function (RTF) system identification

The DTF-GSC beamformer uses the system identification method which exploits the nonstationarity characteristics of the speech signals. In [32], a robust system identification approach adapted to speech signals is proposed. A weighted least-squares optimization criterion is introduced, which considers the uncertainty of the desired signal presence in the observed signals. An asymptotically unbiased estimate for the systems transfer function is derived, and a corresponding recursive online implementation is presented. It has been shown that compared to the nonstationarity-based method, a smaller error variance is achieved and generally shorter observation intervals are required. Furthermore, in the case of a time-varying system, faster convergence and higher reliability of the system identification are obtained by using the proposed method than by using the nonstationarity-based method. Future research may analyze and examine the DTF-GSC performance when the ATFs ratio are estimated using the RTF method, and compare it to the performance obtained using the nonstationarity estimation method.

### 5.2.4 Joint noise reduction and echo cancellation using the ETF-GSC and residual echo cancellation

In Chapter 4, we addressed the problem of joint echo cancellation and noise reduction in a reverberated environment. A solution based on the TF-GSC was presented, denoted ETF-GSC. However, in practice, residual echo still remains

at the output of the ETF-GSC algorithm. This phenomenon is caused by the misadjustment of the adaptive AEC filters and the imposed constraint on filters length. Park *et al.* [33] proposed an integrated acoustic echo and noise canceller, using double-talk detection method and a residual echo cancellation algorithm. In practice, the postprocessors based on noise reduction rarely reduce the residual echo without distorting the near-end talker speech, since the dominant characteristics of the residual echo is also speech. In the residual echo cancellation approach, a linear prediction error filter removes the short-term correlation of the residual echo. The resulting whitened residual echo is cancelled by a noise-reduction filter. Future research may deal with the performance improvement which can be obtained by ETF-GSC followed by residual echo canceller.

# Appendix A

# Proof of Eq. (2.10)

Imposing the constraints defined in (2.6) on (2.9) yields

$$\begin{aligned}
-\mathbf{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\left[\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right] &= \mathcal{F}(e^{j\omega}) \\
-\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\left[\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right] &= 0.
\end{aligned}$$

Solving for the Lagrange multipliers yields:

$$\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix} = C^{-1}\begin{bmatrix}-\mathcal{F}(e^{j\omega})\\ 0\end{bmatrix}$$

where $C$ is defined as

$$C \equiv \begin{bmatrix}\mathbf{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{A}(e^{j\omega}) & \mathbf{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\\ \mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{A}(e^{j\omega}) & \mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\end{bmatrix}.$$

The solution for the linear equations is then given by

$$\lambda_1 = -\alpha^{-1}\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\mathcal{F}(e^{j\omega})$$

$$\lambda_2 = \alpha^{-1}\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})$$

where

$$\begin{aligned}
\alpha \equiv\ & \mathbf{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega}) \\
& \cdot\left[\mathbf{A}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{B}(e^{j\omega}) - \mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\right].
\end{aligned}$$

Therefore the optimal solution is:

$$\mathbf{W}^{opt}(t,e^{j\omega}) = -\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\left(\lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega})\right) = \frac{\mathcal{F}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})}{\mathbf{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})}$$

$$\cdot\frac{\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{A}(e^{j\omega}) - \mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{B}(e^{j\omega})}{\left[\mathbf{A}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{B}(e^{j\omega}) - \mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{A}(e^{j\omega})\right]}.$$

Dividing the nominator and denominator of last term by

$$(\boldsymbol{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega}))(\boldsymbol{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{B}(e^{j\omega}))$$

yields

$$\mathbf{W}^{opt}(t,e^{j\omega}) = \mathcal{F}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})$$
$$\cdot\frac{\frac{\mathbf{A}(\mathbf{e}^{\mathbf{j}\omega})}{\boldsymbol{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})} - \frac{\boldsymbol{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})\boldsymbol{B}(e^{j\omega})}{(\boldsymbol{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega}))(\boldsymbol{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{B}(e^{j\omega}))}}{1 - \frac{\left|\boldsymbol{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega})\right|^2}{(\boldsymbol{A}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{A}(e^{j\omega}))(\boldsymbol{B}^{\dagger}(e^{j\omega})\Phi^{-1}_{\boldsymbol{ZZ}}(t,e^{j\omega})\boldsymbol{B}(e^{j\omega}))}}.$$

Rearranging terms and using definitions (2.11) and (2.12) yields (2.10).

# Appendix B

# Proof of Eq. (2.14)

Imposing our constraints on (2.13) yields the following linear equations:

$$\begin{aligned}\mathbf{A}^{\dagger}(e^{j\omega})\left[\mathbf{W}(t,e^{j\omega})-\mu\left(\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})+\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right)\right]&=\mathcal{F}(e^{j\omega})\\ \mathbf{B}^{\dagger}(e^{j\omega})\left[\mathbf{W}(t,e^{j\omega})-\mu\left(\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})+\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right)\right]&=0.\end{aligned} \tag{B.1}$$

Rearranging terms in (B.1) yields

$$\begin{aligned}&\lambda_1\mu\left\|\mathbf{A}(e^{j\omega})\right\|^2+\lambda_2\mu\mathbf{A}^{\dagger}(e^{j\omega})\mathbf{B}(e^{j\omega})=\\ &\quad\mathbf{A}^{\dagger}(e^{j\omega})\mathbf{W}(t,e^{j\omega})-\mu\mathbf{A}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})-\mathcal{F}(e^{j\omega})\end{aligned}$$

and

$$\begin{aligned}&\lambda_1\mu\mathbf{B}^{\dagger}(e^{j\omega})\mathbf{A}(e^{j\omega})+\lambda_2\mu\left\|\mathbf{B}(e^{j\omega})\right\|^2=\\ &\quad\mathbf{B}^{\dagger}(e^{j\omega})\mathbf{W}(t,e^{j\omega})-\mu\mathbf{B}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\end{aligned}$$

and therefore

$$\begin{aligned}&\begin{bmatrix}\left\|\mathbf{A}(e^{j\omega})\right\|^2 & \mathbf{A}^{\dagger}(e^{j\omega})\mathbf{B}(e^{j\omega})\\ \mathbf{B}^{\dagger}(e^{j\omega})\mathbf{A}(e^{j\omega}) & \left\|\mathbf{B}(e^{j\omega})\right\|^2\end{bmatrix}\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix}=\\ &\quad\frac{1}{\mu}\begin{bmatrix}\mathbf{A}^{\dagger}(e^{j\omega})\mathbf{W}(t,e^{j\omega})-\mu\mathbf{A}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})-\mathcal{F}(e^{j\omega})\\ \mathbf{B}^{\dagger}(e^{j\omega})\mathbf{W}(t,e^{j\omega})-\mu\mathbf{B}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\end{bmatrix}.\end{aligned} \tag{B.2}$$

Solving the linear equations yields:

$$\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix} = \frac{1}{\mu}\ \frac{\begin{bmatrix}\|\mathbf{B}(e^{j\omega})\|^2 & -\mathbf{A}^\dagger(e^{j\omega})\mathbf{B}(e^{j\omega})\\ -\mathbf{B}^\dagger(e^{j\omega})\mathbf{A}(e^{j\omega}) & \|\mathbf{A}(e^{j\omega})\|^2\end{bmatrix}}{\|\mathbf{A}(e^{j\omega})\|^2\|\mathbf{B}(e^{j\omega})\|^2 - \mathbf{A}^\dagger(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\mathbf{A}(e^{j\omega})} \tag{B.3}$$

$$\times\begin{bmatrix}\mathbf{A}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{A}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega})\\ \mathbf{B}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{B}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\end{bmatrix}.$$

Define

$$\alpha \equiv \left\|\mathbf{A}(e^{j\omega})\right\|^2\left\|\mathbf{B}(e^{j\omega})\right\|^2 - \tag{B.4}$$
$$\mathbf{A}^\dagger(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\mathbf{A}(e^{j\omega}) = \left(1-\left\|\rho(e^{j\omega})\right\|^2\right)\left\|\mathbf{A}(e^{j\omega})\right\|^2\left\|\mathbf{B}(e^{j\omega})\right\|^2$$

then

$$\lambda_1 = (\alpha\mu)^{-1}\cdot \tag{B.5}$$
$$\Big[\left\|\mathbf{B}(e^{j\omega})\right\|^2\left(\mathbf{A}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{A}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega})\right)$$
$$-\mathbf{A}^\dagger(e^{j\omega})\mathbf{B}(e^{j\omega})\left(\mathbf{B}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{B}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right)\Big]$$

$$\lambda_2 = (\alpha\mu)^{-1}\cdot \tag{B.6}$$
$$\Big[-\mathbf{B}^\dagger(e^{j\omega})\mathbf{A}(e^{j\omega})\left(\mathbf{A}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{A}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathcal{F}(e^{j\omega})\right)$$
$$+\left\|\mathbf{A}(e^{j\omega})\right\|^2\left(\mathbf{B}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{B}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right)\Big].$$

The next expression is utilized in calculating $\mathbf{W}(t+1,e^{j\omega})$:

$$\mu\left[\lambda_1\mathbf{A}(e^{j\omega}) + \lambda_2\mathbf{B}(e^{j\omega})\right] = \alpha\left\|\mathbf{B}(e^{j\omega})\right\|^2\cdot$$
$$\left[\mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})\right]$$
$$-\alpha\left[\mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega})\right.$$
$$\left.-\mu\mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right]$$
$$-\alpha\left[\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega})\right.$$
$$\left.-\mu\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\mathbf{A}(e^{j\omega})\mathbf{A}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right] + \alpha\left[\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\mathbf{A}(e^{j\omega})\mathcal{F}(e^{j\omega})\right.$$
$$\left.+\left\|\mathbf{A}(e^{j\omega})\right\|^2\left(\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\mathbf{W}(t,e^{j\omega}) - \mu\mathbf{B}(e^{j\omega})\mathbf{B}^\dagger(e^{j\omega})\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\right)\right]. \tag{B.7}$$

Rearranging terms in (B.7) yields

$$\begin{aligned}\mu\left[\lambda_1\mathbf{A}(e^{j\omega})+\lambda_2\mathbf{B}(e^{j\omega})\right] &= \alpha\left[\mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})\left(\left\|\mathbf{B}(e^{j\omega})\right\|^2\mathbf{I}-\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\right)\right.\\ &\left.-\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\left(\mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})-\left\|\mathbf{A}(e^{j\omega})\right\|^2\mathbf{I}\right)\right]\mathbf{W}(t,e^{j\omega})\\ &-\alpha\mu\left[\mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})\left(\left\|\mathbf{B}(e^{j\omega})\right\|^2\mathbf{I}-\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\right)\right.\\ &\left.-\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\left(\mathbf{A}(e^{j\omega})\mathbf{A}^{\dagger}(e^{j\omega})-\left\|\mathbf{A}(e^{j\omega})\right\|^2\mathbf{I}\right)\right]\cdot\Phi_{\boldsymbol{ZZ}}(t,e^{j\omega})\mathbf{W}(t,e^{j\omega})\\ &+\alpha\left[-\left\|\mathbf{B}(e^{j\omega})\right\|^2\mathbf{A}(e^{j\omega})+\mathbf{B}(e^{j\omega})\mathbf{B}^{\dagger}(e^{j\omega})\mathbf{A}(e^{j\omega})\right]\mathcal{F}(e^{j\omega}).\end{aligned} \tag{B.8}$$

Substituting (B.8) into (2.13) we obtain (2.14).

# Appendix C

# Derivation of expected noise reduction for coherent noise field

Recall (3.10) and define,

$$\mathcal{X}(t,e^{j\omega}) = \Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\left(\widehat{\mathcal{H}}(e^{j\omega})^{\dagger}\Phi_{\boldsymbol{NN}}(t,e^{j\omega})\widehat{\mathcal{H}}(e^{j\omega})\right)^{-1}\widehat{\mathcal{H}}^{\dagger}(e^{j\omega})\Phi_{\boldsymbol{NN}}(t,e^{j\omega})$$

For clarity and simplicity we will omit the time and frequency dependence in the next derivation. Thus,

$$\mathcal{X}(t,e^{j\omega}) = \mathcal{X} = \Phi_{\boldsymbol{NN}}\widehat{\mathcal{H}}\left(\widehat{\mathcal{H}}^{\dagger}\Phi_{\boldsymbol{NN}}\widehat{\mathcal{H}}\right)^{-1}\widehat{\mathcal{H}}^{\dagger}\Phi_{\boldsymbol{NN}}.$$

We further denote,

$$\mathcal{X} = \mathcal{K}\times\mathcal{L}\times\mathcal{M}$$

where, $\mathcal{K} = \Phi_{\boldsymbol{NN}}\widehat{\mathcal{H}}$, $\mathcal{L} = \left(\widehat{\mathcal{H}}^{\dagger}\Phi_{\boldsymbol{NN}}\widehat{\mathcal{H}}\right)^{-1}$ and $\mathcal{M} = \widehat{\mathcal{H}}^{\dagger}\Phi_{\boldsymbol{NN}}$. Thus, $\mathcal{X}$ is a multiplication of three terms. Starting from $\mathcal{L}$ and using the detailed noise structure,

$$\mathcal{L} = \left(\widehat{\mathcal{H}}^{\dagger}(\Phi_{nn}\boldsymbol{D}\boldsymbol{D}^{\dagger}+\varepsilon I)\widehat{\mathcal{H}}\right)^{-1} = \left(\Phi_{nn}(\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D})(\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D})^{\dagger}+\varepsilon\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}}\right)^{-1}.$$

If $\boldsymbol{D}(e^{j\omega}) = \boldsymbol{A}(e^{j\omega})$, i.e., the noise source is located exactly at the desired signal location, than $\mathcal{H}^{\dagger}(e^{j\omega})\boldsymbol{D}(e^{j\omega}) = 0$, and the calculation of the inverse is straight forward, yielding $\mathcal{L} = (\varepsilon\widehat{\mathcal{H}}^{\dagger}\mathcal{H})^{-1}$, $\mathcal{K} = \varepsilon\widehat{\mathcal{H}}$ and $\mathcal{M} = \varepsilon\widehat{\mathcal{H}}^{\dagger}$. Collecting terms we obtain, $\mathcal{X} = \frac{\varepsilon^2}{\varepsilon}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\mathcal{H})^{-1}\widehat{\mathcal{H}}^{\dagger} \overset{\varepsilon\to 0}{\longrightarrow} 0$. Thus, no extra noise reduction is produced by the noise canceller branch, as expected. The total noise part of the output is

given by,

$$\Phi_{yy}^{n}(t,e^{j\omega}) = \Phi_{fbf}^{n}(t,e^{j\omega}) = \Phi_{nn}(t,e^{j\omega})|\mathcal{F}(e^{j\omega})|^2|A_1(e^{j\omega})|^2$$

which is exactly the signal part derived in Section 3.4.

For the general case, $\boldsymbol{D}(e^{j\omega}) \neq \boldsymbol{A}(e^{j\omega})$, we use the *Matrix Inversion Lemma*, yielding:

$$\mathcal{L} = \left( \frac{1}{\varepsilon}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1} - \frac{\frac{1}{\varepsilon^2}\Phi_{nn}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}}{1+\frac{1}{\varepsilon}\Phi_{nn}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}} \right).$$

Now, using the approximation $\frac{1}{1+\mu} \approx 1-\mu$, for $\mu \to 0$ ($\mu$ properly defined), yields,

$$\mathcal{L} = \left( \frac{1}{\varepsilon}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1} - \right.$$
$$\left. \frac{\frac{1}{\varepsilon}\Phi_{nn}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}}{\Phi_{nn}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}} + \frac{\Phi_{nn}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}}{\left(\Phi_{nn}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}\right)^2} \right).$$

Now, calculating $\mathcal{X}$,

$$\begin{aligned}\mathcal{X} &= \mathcal{K}\mathcal{L}\mathcal{M} = \Phi_{\boldsymbol{NN}}\widehat{\mathcal{H}}\mathcal{L}\widehat{\mathcal{H}}^{\dagger}\Phi_{\boldsymbol{NN}} = \\ &(\Phi nn\boldsymbol{D}\boldsymbol{D}^{\dagger}+\varepsilon I)\widehat{\mathcal{H}}\mathcal{L}\widehat{\mathcal{H}}^{\dagger}(\Phi nn\boldsymbol{D}\boldsymbol{D}^{\dagger}+\varepsilon I) \\ &\stackrel{\triangle}{=} (\mathcal{K}_1+\mathcal{K}_2)\widehat{\mathcal{H}}(\mathcal{L}_1+\mathcal{L}_2+\mathcal{L}_3)\widehat{\mathcal{H}}^{\dagger}(\mathcal{M}_1+\mathcal{M}_2).\end{aligned}$$

with the obvious definitions of $\mathcal{K}_1$, $\mathcal{K}_2$, $\mathcal{L}_1$, $\mathcal{L}_2$, $\mathcal{L}_3$, $\mathcal{M}_1$, $\mathcal{M}_2$. Opening the brackets we have twelve terms:

I. $\mathcal{K}_1\widehat{\mathcal{H}}\mathcal{L}_1\widehat{\mathcal{H}}^{\dagger}\mathcal{M}_1 = \frac{1}{\varepsilon}\Phi_{nn}^2\boldsymbol{D}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}\boldsymbol{D}^{\dagger}$

II. $\mathcal{K}_1\widehat{\mathcal{H}}\mathcal{L}_2\widehat{\mathcal{H}}^{\dagger}\mathcal{M}_1 = -\frac{1}{\varepsilon}\Phi_{nn}^3 \times \frac{\boldsymbol{D}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\boldsymbol{D}\boldsymbol{D}^{\dagger}}{\Phi_{nn}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}} =$
$-\frac{1}{\varepsilon}\Phi_{nn}^2\boldsymbol{D}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D}\boldsymbol{D}^{\dagger}$

III. $\mathcal{K}_1\widehat{\mathcal{H}}\mathcal{L}_3\widehat{\mathcal{H}}^{\dagger}\mathcal{M}_1 = \frac{\Phi_{nn}^3\boldsymbol{D}(\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D})^2\boldsymbol{D}^{\dagger}}{(\Phi_{nn}\boldsymbol{D}^{\dagger}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^{\dagger}\boldsymbol{D})^2} = \Phi_{nn}\boldsymbol{D}\boldsymbol{D}^{\dagger}$

IV. $\mathcal{K}_2\widehat{\mathcal{H}}\mathcal{L}_1\widehat{\mathcal{H}}^{\dagger}\mathcal{M}_1 = \Phi_{nn}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^{\dagger}\widehat{\mathcal{H}})^{-1}\mathcal{H}^{\dagger}\boldsymbol{D}\boldsymbol{D}^{\dagger}$

V. $\mathcal{K}_2\widehat{\mathcal{H}}\mathcal{L}_2\widehat{\mathcal{H}}^\dagger\mathcal{M}_1 = -\Phi_{nn}^2\frac{\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\boldsymbol{D}\boldsymbol{D}^\dagger}{\Phi_{nn}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}} =$
$-\Phi_{nn}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger$

VI. $\mathcal{K}_2\widehat{\mathcal{H}}\mathcal{L}_3\widehat{\mathcal{H}}^\dagger\mathcal{M}_1 = \varepsilon\frac{\Phi_{nn}^2\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger}{(\Phi_{nn}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D})^2} = \varepsilon\frac{\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}}^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger}{\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}}$

VII. $\mathcal{K}_1\widehat{\mathcal{H}}\mathcal{L}_1\widehat{\mathcal{H}}^\dagger\mathcal{M}_2 = \Phi_{nn}\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger$

VIII. $\mathcal{K}_1\widehat{\mathcal{H}}\mathcal{L}_2\widehat{\mathcal{H}}^\dagger\mathcal{M}_2 = -\Phi_{nn}^2\frac{\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger}{\Phi_{nn}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}} =$
$-\Phi_{nn}\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger$

IX. $\mathcal{K}_1\widehat{\mathcal{H}}\mathcal{L}_3\widehat{\mathcal{H}}^\dagger\mathcal{M}_2 = \varepsilon\frac{\Phi_{nn}^2\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger}{(\Phi_{nn}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D})^2} = \varepsilon\frac{\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger}{\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}}$

X. $\mathcal{K}_2\widehat{\mathcal{H}}\mathcal{L}_1\widehat{\mathcal{H}}^\dagger\mathcal{M}_2 = \varepsilon\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger$

XI. $\mathcal{K}_2\widehat{\mathcal{H}}\mathcal{L}_2\widehat{\mathcal{H}}^\dagger\mathcal{M}_2 = -\varepsilon\frac{\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger}{\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}}$

XII. $\mathcal{K}_2\widehat{\mathcal{H}}\mathcal{L}_3\widehat{\mathcal{H}}^\dagger\mathcal{M}_2 = \varepsilon^2\frac{\Phi_{nn}\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger}{(\Phi_{nn}\boldsymbol{D}^\dagger\widehat{\mathcal{H}}(\widehat{\mathcal{H}}^\dagger\widehat{\mathcal{H}})^{-1}\widehat{\mathcal{H}}^\dagger\boldsymbol{D})^2}$

Note, that terms I,II terms IV,V and terms VII,VIII eliminate each other, and that terms VI,IX,X,XI,XII tend to zero as $\varepsilon$ tends to zero. Only term III is left. Collecting all terms, we have for the noise part of the output,

$$\Phi_{yy}^n(t, e^{j\omega}) =$$
$$\boldsymbol{W_0}^\dagger(e^{j\omega})\left(\Phi_{\boldsymbol{NN}}(t, e^{j\omega})\boldsymbol{H}(e^{j\omega}) - \boldsymbol{H}^\dagger(e^{j\omega})\Phi_{nn}(t, e^{j\omega})\boldsymbol{D}(e^{j\omega})\boldsymbol{D}^\dagger(e^{j\omega})\right)\boldsymbol{W_0}(e^{j\omega}) = 0$$

# Bibliography


[1] O.L. Frost III, "An Algorithm for Linearly Constrained Adaptive Array Processing," *Proceedings of the IEEE*, vol. 60, no. 8, pp. 926–935, Jan. 1972.

[2] L. J. Griffiths and C. W. Jim, "An Alternative Approach to Linearly Constrained Adaptive Beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[3] B.D. Van Veen and K.M. Buckley, "Beamforming: A Versatile Approach to Spatial Filtering," *IEEE Acoustics, Speech and Signal Proc. magazine*, pp. 4–24, Apr. 1988.

[4] S. Affes and Y. Grenier, "A Signal Subspace Tracking Algorithm for Microphone Array Processing of Speech," *IEEE transactions on Speech and Audio Processing*, vol. 5, no. 5, pp. 425–437, Sep. 1997.

[5] S. Gannot, D. Burshtein, and E. Weinstein, "Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech," *IEEE Trans. Signal Processing*, vol. 49, no. 8, pp. 1614–1626, August 2001.

[6] L.C. Parra and C. Spence, "Convolutive Blind Seperation of Non-Stationary Sources," *IEEE Trans. on Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[7] L. C. Parra and C. V. Alvino, "Geometric Source Separation: Merging Convolutive Source Separation With Geometric Beamforming," *IEEE Trans. on Speech and Audio Processing*, vol. 10, no. 6, pp. 352–362, Sep. 2002.

[8] S. Kurita, H. Saruwatari, S. Kajita, K. Takeda, and F. Itakura, "Evaluation of Blind Signal Separation method using Directivity Pattern under Reverberant Conditions," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Istanbul, Turkey, Jun. 2000, IEEE, pp. 3140–3143.

[9] M. Ikram and D. Morgan, "A Beamformer Approach to Permutation Alignment for Multichannel Frequency-Domain Blind Source Separation," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Orlando, USA, May 2002, IEEE, pp. 881–884.

[10] M. Knaak, S. Araki, and S. Makino, "Geometrically Constrained ICA for Robust Separation of Sound Mixtures," in *Proc. 4th International Symposium on ICA and BSS*, Nara, Japan, Apr. 2003, IEEE, pp. 951–956.

[11] A. Hyvärinen, J. Karhunen, and E. Oja, *Independent Component Analysis*, Adaptive and Learning Systems. Wiley, 1st edition, 2001.

[12] G. Reuven, S. Gannot, and I. Cohen, "Joint Acoustic Echo Cancellation and Transfer Function GSC in the Frequency Domain," in *IEEEI*, Herzlia, Israel, Sept. 2004, pp. 412–415.

[13] S. Affes and Y. Grenier, "A Source Subspace Tracking Array of Microphones for Double Talk Situations," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Munich, Germany, Apr. 1996, pp. 269–272.

[14] W. Kellermann, "Strategies for Combining Acoustic Echo Cancellation and Adaptive Beamforming Microphone Arrays," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Munich, Apr. 1997, pp. 219–222.

[15] S. Doclo, M. Moonen, and E.D. Clippel, "Combined acoustic echo and noise reduction using gsvd-based optimal filtering," in *IEEE Int. Conf. Acoust. Speech and Sig. Proc. (ICASSP)*, Istanbul, Turkey, June 2000, pp. 1061–1064.

[16] G. Rombouts and M. Moonen, "An Integrated Approach to Acoustic Noise and Echo Cancellation," *Signal Processing*, vol. 85, no. 4, pp. 849–871, Apr. 2005.

[17] G. Reuven S. Gannot and I. Cohen, "Dual Source TF-GSC and its Application to Echo Cancellation," in *The International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Eindhoven, The Netherlands, Sep. 2005, pp. 27–30.

[18] G. Reuven, S. Gannot, and I. Cohen, "Audio Sample Files," `http://www.eng.biu.ac.il/~gannot`, Nov. 2005.

[19] D.H. Brandwood, "A Complex Gradient Operator and its Application in Adaptive Array Theory," *IEE Proceedings*, vol. 130, no. 1, Parts F and H, pp. 11–16, Feb. 1983.

[20] G. Strang, *Linear Algebra and its Application*, Academic press, 2nd edition edition, 1980.

[21] B. Widrow, J.R. Glover Jr., J.M. McCool, J. Kaunitz, C.S. Williams, R.H. Hearn, J.R. Zeider, E. Dong Jr., and R.C. Goodlin, "Adaptive Noise Cancelling: Principals and Applications," *Proceeding of the IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[22] S. Nordholm, I. Claesson, and P. Eriksson, "The Broadband Wiener solution for Griffiths-Jim Beamformers," *IEEE trans. on Signal Proc.*, vol. 40, no. 2, pp. 474–478, Feb. 1992.

[23] J. Bitzer, K.-D. Kammeyer, and K.U. Simmer, "An Alternative Implementation of the Superdirective Beamformer," in *IEEE Workshop on Application of Signal Processing to Audio and Acoustics*, New Paltz, New York, USA, Oct. 1999.

[24] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, "Acoustic-phonetic continuous speech corpus (timit)," CD-ROM, Oct. 1991.

[25] A. Varga and H.J.M. Steeneken, "Assessment for Automatic Speech Recognition: II. NOISEX-92: A Database and an Experiment to Study the Effect of Additive Noise on Speech Recognition Systems," *Speech Communication*, vol. 12, pp. 247–251, Jul. 1993.

[26] J.B. Allen and D.A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," *J. Acoust. Soc. Am*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[27] N. Dal-Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio application," *Signal Processing*, vol. 15, no. 4, pp. 43–56, Jul. 1988.

[28] S. Gannot, D. Burshtein, and E. Weinstein, "Analysis of the Power Spectral Deviation of the General Transfer Function GSC," *IEEE Trans. Signal Processing*, vol. 52, no. 4, pp. 1115–1121, April 2004.

[29] J. Bitzer, K.U. Simmer, and K.D. Kammeyer, "Theoretical Noise Reduction Limits of the Generalized Sidelobe Canceller (GSC) for Speech Enhancement," in *Proc. Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, Phoenix, Arizona, USA, Mar. 1999, IEEE, pp. 2965–2968.

[30] W. Kellermann, "Joint Design of Acoustic Echo Cancellation and Adaptive Beamforming for Microphone Arrays," in *The International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Imperial College, London, UK, 1997, pp. 81–84.

[31] W. Herbordt and W. Kellermann, "GSAEC - Acoustic echo cancellation embedded into the generalized sidelobe canceller," in *Proc. European Signal*

*Processing Conference ( EUSIPCO)*, Tampere, Finland, Sept. 2000, vol. 3, p. 18431846.

[32] I. Cohen, "Relative Transfer Function Identification Using Speech Signals," *IEEE Trans. on Speech and Audio Processing*, vol. 12, no. 5, pp. 451–459, Nov. 2004.

[33] C. Lee S.J. Park, C.G. Cho and D.H. Youn, "Integrated Echo and Noise Canceler for Hands-Free Applications," *IEEE TRANS. ON CIRCUITS AND SYSTEMSII: ANALOG AND DIGITAL SIGNAL PROCESSING*, vol. 49, no. 3, pp. 188–195, March 2002.