Distributed Speech Processing for Microphone Networks

Shmulik Markovich-Golan

Faculty of Engineering

Ph.D. Thesis

Submitted to the Senate of Bar-Ilan University

Ramat-Gan, Israel March 2013

This work was carried out under the supervision of:

prof. Sharon Gannot, the Faculty of Engineering at Bar-Ilan University

prof. Israel Cohen, the Department of Electrical Eng. at the Technion

Acknowledgement

I wish to express my deep gratitude and appreciation to my supervisors Prof. Sharon Gannot and Prof. Israel Cohen for their guidance and dedicated supervision. Thank you for your professional support, for your encouragement to perfection, and for many valuable suggestions throughout all stages of this research.

The generous financial help of the President of Bar-Ilan is gratefully acknowledged.

Thanks to my mother Orit, my late father Yaakov, to Shabtai Aharon and to Eli and Orly Golan. Special thanks to my beloved Liran who encouraged and supported me through the whole way.

Contents

List of Figures

List of Tables

List of Algorithms

Notation

x	Scalar
$\boldsymbol{x}$	Column vector
x_i	The ith element of the vector $\boldsymbol{x}$
$\boldsymbol{A}$	Matrix
A_{ij}	The (i, j) element of the matrix $\boldsymbol{A}$
$\boldsymbol{A}^{-1}$	Matrix inverse
$(\cdot)^T$	Transpose operation
$(\cdot)^*$	Conjugate operation
$(\cdot)^H$	Transpose-conjugate operation
$\mathrm{diag}\{\boldsymbol{x}\}$	Diagonal matrix with the vector $\boldsymbol{x}$ on its diagonal
$(\cdot)^{\frac{1}{2}}$	For diagonal matrices, a diagonal matrix with the square root of the diagonal
$\|\cdot\|$	Euclidian norm operation
$\boldsymbol{I}$	Identity matrix
$\mathrm{E}\,(\cdot)$	Expectation operation
$x(\ell, k)$	Time-frequency coefficient

Abbreviations

ADC analog to digital converter

ANC adaptive noise canceler

AOA angle of arrival

AR auto regressive

ATF acoustic transfer function

BF beamformer

BM blocking matrix

B-MWF binaural multichannel Wiener filter

BSS blind source separation

c.d.f. cumulative distribution function

CB collaborative beamforming

CLT central limit theorem

CUD constraint update decremental

CUD-GSC constraint update decremental (CUD) generalized sidelobe canceler (GSC)

CUD-LCMV CUD linearly constrained minimum variance (LCMV)

CUI constraint update incremental

CUI-GSC constraint update incremental (CUI) GSC

CUI-LCMV CUI LCMV

DAC digital to analog converter

DANSE distributed adaptive node-specific signal estimation

DB-MWF iterative distributed speech distortion weighted multichannel Wiener filter (SDW-MWF)

DFT discrete Fourier transform

DGSC distributed GSC

DOA direction of arrival

DRR direct to reverberant ratio

DS delay and sum

DS-BF delay and sum beamformer

DS-GSC distributed single-constraint generalized sidelobe canceler

DTF-GSC dual source transfer function generalized sidelobe canceler

EVD eigenvalue decomposition

FBF fixed beamformer

FIR finite impulse response

GEVD generalized eigenvalue decomposition

GSC generalized sidelobe canceler

GSVD generalized singular value decomposition

i.i.d. independent identically distributed

ICA independent component analysis

ILD interaural level difference

INR interference to noise ratio

ITD interaural time difference

LC linearly constrained

LC-DANSE linearly constrained distributed adaptive node-specific signal estimation (DANSE)

LCMV linearly constrained minimum variance

LMS least mean squares

MEMS micro electro-mechanic systems

MFA matched filter array

MIMO multiple input multiple output

MIPS mega instruction per second

MMSE minimum mean square error

MSDW-MWF multiple speech distortions weighted multichannel Wiener filter

MSE mean squared error

MSNR maximum signal to noise ratio

MUSIC multiple signal classification

MVDR minimum variance distortionless response

MWF multichannel Wiener filter

NC noise canceler

NLMS normalized least mean squares (LMS)

NR noise reduction

p.d.f. probability distribution function

PASTd projection approximation subspace tracking deflation

PSD power spectral density

QRD orthogonal triangular decomposition

RBS reference-broadcast synchronization

RIR room impulse response

RTF relative transfer function

RV random variable

SDR signal to distortion ratio

SDW speech distortion weighted

SDW-MWF speech distortion weighted multichannel Wiener filter

SF-GSC straightforward GSC

SF-LCMV straightforward LCMV

SIR signal to interference ratio

SNR signal to noise ratio

SOI signal of interest

STFT short time Fourier transform

SUD sensor update decremental

SUD-GSC sensor update decremental (SUD) GSC

SUD-LCMV SUD LCMV

SUI sensor update incremental

SUI-GSC sensor update incremental (SUI) GSC

SUI-LCMV SUI LCMV

SVD singular value decomposition

TDOA time difference of arrival

TF transfer function

TF-GSC transfer function generalized sidelobe canceler

TRINICON triple-N independent component analysis (ICA) for convolutive mixtures

VAD voice activity detector

WASN wireless acoustic sensor network

WSN wireless sensor network

WSS wide-sense stationary

Abstract

The concept of distributed sensor networks is becoming more realistic as technology advances in the fields of nano-technology, micro electro-mechanic systems (MEMS) and communication. A distributed sensor network comprises scattered *nodes* which are autonomous, self-powered modules consisting of sensors, actuators and communication capabilities. Their layout and connectivity graph are usually random and dynamic. Distributed sensor networks have a broad range of applications which can be categorized in ecology, environment monitoring, medical, security and surveillance. In this dissertation we develop algorithms for distributed sensor networks with applications to speech processing. Such wireless acoustic sensor networks (WASNs) can be found useful in the following scenarios for example: 1) Ambient immersive communications: nowadays, almost everyone carries 'his/her personal microphones' as part of the cellular phone, laptop computer or tablet, these spatially distributed sensors allow exploitation of spatial information in addition to spectro-temporal information. These sensors make the establishment of an *ad hoc* (distributed) microphone network feasible and allow the application of sophisticated signal extraction algorithms without the need to pre-install expensive audio systems; 2) Smart homes: intelligent networks of microphones are crucial components for control and monitoring systems as well as for communication in emergency cases; 3) Law Enforcement: Law enforcement authorities, e.g., police and homeland security, use eavesdropping and acoustic surveillance of public spaces as part of their regular procedure. This is usually done under adverse conditions.

The availability of only partial information in the nodes, the dynamics of the network, and the limited communication, connectivity and power capabilities call upon developing novel algorithms that address these challenges. The latter challenges are typical to distributed algorithms and cannot be found in classical array processing algorithms.

The contribution of the dissertation is threefold. First, distributed versions of classical beamforming algorithms are developed. Specifically, the LCMV beamformer with multiple

speakers scenario is considered, and a time-recursive distributed version of it is derived. Second, a novel framework for evaluating beamformers in WASN is derived. In classical array processing the layout of the array is usually pre-determined to fit the problem at hand, contrary to WASN where the layout can be random and dynamic. The performance of data-dependent beamformers (BFs) with randomly located microphones is analyzed. The analysis may serve as a design guideline for determining the number of nodes/microphones required in order to meet a desired performance level. Third, we address the problem of reducing the complexity of applying distributed BFs, without sacrificing performance. The dynamic nature of the sources, the environment and the network connectivity, require frequent modification to the applied BF. Efficient methods for modifying the BF in accordance with this dynamics are developed.

Chapter 1

Introduction

Speech enhancement techniques, utilizing microphone arrays, have attracted the attention of many researchers for the last thirty years, especially in hands-free communication tasks. Typical problems in this field are: noise reduction, echo cancelation, speaker extraction and de-reverberation. Beamforming algorithms extend the dimension of the solutions and introduces spatial filtering in addition to the classical temporal-spectral filtering. Usually, the received speech signals are contaminated by interfering sources, such as competing speakers and noise sources, and also distorted by the reverberating environment. Whereas single microphone algorithms might show satisfactory results in noise reduction, they are not very-well suited in competing speaker mitigation task, as they lack the spatial information, or the statistical diversity used by multi-microphone algorithms. Despite the obvious advantages over single-microphone systems, traditional microphone arrays still suffer from severe performance limitations.

The relatively small aperture of conventional arrays is a limiting factor in the performance of spatial processing algorithms, since they only sample the sound field locally, typically at a relatively large distance from the target source(s). In these scenarios, low signal to noise ratio (SNR) and low direct to reverberant ratio (DRR) are expected, resulting in deteriorated performance of regular 'condensed' small-aperture microphone arrays. Consequently, classical condensed microphone-arrays cannot serve as a complete solution in the following example scenarios: 1) *Immersive communications*- state of the art telecommunication systems attempt not only to faithfully convey the semantics of a conversation, but also to enable natural experiences and interactions among physically separated people, as if they shared the same room. Many telephone conversations, especially those held in hands-free mode, are corrupted

with background noise, interfering signals and room reverberation; 2) *Smart houses* denotes a unified system for controlling all systems of the house (lights, air-condition, electronic devices such as television, sound system etc.). In this aspect, intelligent networks of microphones are crucial components for control and monitor systems as well as for communication in emergency cases. 3) *Law enforcement* authorities, e.g., police and homeland security, use eavesdropping and acoustic surveillance of public spaces as part of their regular procedure. This is usually done under adverse conditions. The microphones should be deployed in a large area to ensure proper reception of the desired speakers.

A straightforward *centralized* approach is to place sensors in the vast environment and to convey all available data from them to a fusion centre where the processing is performed. Though optimal, this simple method requires transmission of huge amounts of data. Moreover, the aforementioned simple algorithm is sensitive to a failure of the fusion centre, rendering the sensor network useless. Another disadvantage caused by the structure of the centralized solution, is the long communication link between sensors and the fusion centre, which might be comprised of several hops (when the fusion centre and the sensor cannot communicate directly) manifested as slow adaptation to the dynamically changing network or environment [1, 2].

Recent technological advances in the fields of nano-technology, MEMS, together with improved communication capabilities, have made the vision of a distributed sensor network feasible. A wireless sensor network (WSN) comprises several nodes (WSN modules) interconnected in some manner via a wireless medium. Each node consists of one or more sensors, a processing unit and a wireless communication module allowing them to exchange data. The goal of the system is to perceive some physical phenomenon, to process it, and to yield a required result. In classical array processing systems, the sensing and the processing of the acquired data are concentrated in a single location denoted a *fusion center*. A phenomenon originating in the enclosure, results in a disturbance that propagates in space. The closer the sensors are to the origin of the phenomenon, the higher is the SNR of the acquired signal, resulting in lower estimation errors and better quality at the output of the signal processing procedure.

The concept of the WSN is to divide the system resources (sensors, processing units and actuators) among the nodes and to provide a scalable, fully covering the environment, easy to deploy, and robust structure. The wireless interface allows for the extension of the sensing range beyond the limits of the wired fusion center systems. The distribution of the sensors

in a larger volume enables better coverage with higher SNR, higher DRR and less sensitivity to sources locations. For a survey on the topic of WSN please refer to [1, 3, 4, 5].

The emerging technology introduces new challenges. 1) **Communication constraints**: Dynamic network topology with limited connectivity and short battery life requires further research in developing efficient, robust and scalable wireless communication methods and protocols (will not be addressed in this work). 2) **Algorithmic**: classical signal processing algorithms are usually designed for the centralized scenarios, where all the sensors' data are available for processing in one centre. Novel distributed algorithms, that involve new optimization criteria, taking the limitations of the problem at hand into consideration, are therefore essential.

This work focuses on the second challenge, namely it aims at developing distributed signal processing algorithms. The derived algorithms should maintain low complexity and communication bandwidth. However, performance issues should not be sacrificed (although, sub-optimal algorithms can be applied). Another requirement introduced by the nature of the problem at hand is the need for robust algorithms, which are not sensitive to failure of few nodes, and which can handle a dynamically changing connectivity of the network. Naturally, most applications require algorithms which also adapt to changes in the environment, or the objects under observation.

Communication (especially in wireless networks) is by far the most energy consuming operation of a node. An alternative signal processing paradigm to the *centralized* processing is the *local* processing in which each node utilizes only its own measurement data, independent of other nodes, rendering communication unnecessary. Although minimizing communication load, this method obviously imposes performance limitations, as only small subset of the data is used in each processing unit. Common systems utilize compression schemes for reducing the required bandwidth for conveying sensor data to the fusion centre. Though straightforward, these methods do not consider the signal processing algorithm that takes place in the fusion centre. Compression might destroy coherence between sensor signals, rendering all beamforming methods useless. Distributed algorithms aim at achieving the performance of the fusion centre paradigm, while considerably reducing the required communication bandwidth. Each node performs a local calculation/filtering and distributes the results in the network. By iteratively merging the local sensors signals with the distributed data, distributed algorithms converge to their centralized version counterparts. Note that, iterations can be performed on a batch of recorded samples, or time-recursively.

In classical beamforming the geometry of the microphone array is usually pre-determined to fit the problem at hand. However, in some distributed sensor network scenarios it is not possible to determine the array's layout. Since, beamformer performance measures depend on geometrical properties of the microphone array, standard performance analysis is inappropriate in WSN. Lo [6] propose to incorporating statistical models into the sensors' spatial distribution. Consequently, he derived a statistical model for the beamformer performance, and obtained better understanding of the array properties (e.g. directivity, beam-width and sidelobe level) of a simple delay and sum (DS) beamformer. However, the analysis was limited to narrow-band sources in non-reflecting environment, and for simple data-independent BFs.

Designing distributed algorithms for WSN necessitates the application of additional considerations. The computational power available at the nodes is limited. Consequently, the amount of calculations for constructing the BF as well as its application should be limited. Furthermore, as each node uses independent clock sources, sampling frequency offsets between nodes are inevitable. The latter results in non-equal sampling rates in the network, and eventually performance degradation.

The contribution of this dissertation is threefold: First, we propose distributed speech processing algorithms based on the LCMV criterion. Second, we analyze beamformer performance in a reverberant environment with a probabilistic nodes layout. Third, we propose methods for coping with the limitation of computational power. Explicitly, we derive an efficient implementation for the blocking matrix (BM) (a computationally demanding part of the BF that will be explained later), and procedures for adjusting the BF to changes in the environment.

The structure of the introduction follows. In Sec. 1.1 classical beamforming concepts are presented. Sec. 1.2 is dedicated to survey of distributed beamforming algorithms. The expected performance of beamforming with randomly located sensors is explored in Sec. 1.3. The computational complexity of applying beamforming algorithms in WSN is discussed in Sec. 1.4.

1.1 Centralized beamformers for Speech processing

Speech enhancement techniques, utilizing microphone arrays, have attracted the attention of many researchers for the last quarter century, especially in hands-free communication tasks.

Usually, the received speech signals are contaminated by interfering sources, such as competing speakers and noise sources, and also distorted by the reverberating environment. Whereas single microphone algorithms might show satisfactory results in noise reduction, they are less suitable in competing speaker mitigation task, as they lack the spatial information, or the statistical diversity used by multi-microphone algorithms.

In this dissertation we address the problem of extracting several desired sources in a reverberant environment comprising of both non-stationary (competing speakers) and stationary interferences. In some cases we will consider the more simple scenario of a single desired speaker.

Two families of microphone array algorithms can be defined, namely, the blind source separation (BSS) family and the beamforming family. BSS aims at separating all the involved sources, regardless of their attribution to the desired or interfering sources [7]. On the other hand, the beamforming family of algorithms, focus on enhancing the sum of the desired sources while treating all other signals as interfering sources. The BSS family of algorithms exploit the independence of the involved sources. ICA algorithms [8, 9] are commonly applied for solving the BSS problem. Independent component analysis (ICA) [10] algorithms are distinguished by the way the source independence is imposed. Commonly used techniques include *second-order statistics* [11], *high-order statistics* [12], and *Information theoretic* based measures [13]. BSS methods can also be used in reverberant environments, but they tend to get very complicated (for time domain approaches [14]) or have an inherent problem of *permutation and gain ambiguity* [15] (for frequency domain algorithms [9]). Souden and Liu [16] propose an interesting combined BSS and linear echo cancellation scheme. They utilized a BSS algorithm, based on higher order statistics, in order to reject non-linear components of the echo, resulting from the speaker. BSS separation problems can be categorized as either *over-determined*, where the number of sources is less then or equal the number of microphones, or *under-determined*, where there are more sources than microphones. Thus far, we discussed solutions for the over-determined problem. Considering the under-determined problem and utilizing the frequency sparsity of speech signals, Kameoka et al. [17] propose a method for separating speech signals in the frequency domain. Assuming non-reverberant environment they resolved the permutation ambiguity based on estimation of the sources' direction of arrival (DOA). BSS methods of acoustic signals can be also utilized for separating musical instruments from a monaural recording [18].

Buchner et al., in a series of contributions, propose the triple-N ICA for convolutive

mixtures (TRINICON), a generic framework for broadband adaptive multiple input multiple output (MIMO) filtering [19, 20, 21, 22, 23]. Specifically, the TRINICON framework is applied successfully in array processing for speech capture. They consider the *determined* or *overdetermined* problem of extracting P sources received as a summation of convulative mixtures by M sensors, where $P \leq M$. The TRINICON brings together various blind and supervised MIMO adaptation techniques that were treated largely independent in the literature. A solution for the problem at hand is comprised of the following general signal processing stages: 1) signal separation: obtain convulative mixtures of each of the sources separately; 2) deconvolution/derverberation: obtain the sources up to a delay and a scaling factor; 3) identification of the mixing system. The proposed solution utilizes fundamental source signal properties: 1) nongaussianity is exploited by using higher-order statistics for ICA; 2) nonwhiteness is exploited by simultaneous minimization of output cross-relations over multiple time-lags; 3) nonstationarity is exploited by simultaneous minimization of output cross-relations at different time-instants. TRINICON provides a versatile tool to the design of adaptive systems, and with its "top-down approach", it has led to various recent advances in the field.

Our research focuses on beamforming algorithms. The term beamforming refers to the design of a spatial-temporal filter. Broadband arrays comprise a set of filters, applied to each received microphone signal, followed by a summation operation. The main objective of the beamformer is to extract a desired signal, impinging on the array from a specific position, from noisy measurements thereof. The simplest structure is the *delay-and-sum* beamformer, which first compensates for the relative delay between distinct microphone signals and then sums the steered signal to form a single output. This beamformer, which is still widely used, can be very effective in mitigating noncoherent, i.e., spatially white noise sources, provided that the number of microphones is relatively high. However, if the noise source is coherent, the noise reduction (NR) is strongly dependent on the direction of arrival of the noise signal. Consequently, the performance of the delay-and-sum beamformer in reverberant environments is often insufficient. Jan and Flanagan [24] extended the delay-and-sum concept by introducing the so called *filter-and-sum* beamformer. This structure, designed for multipath environments, namely reverberant enclosures, replaces the simpler delay compensator with a matched filter. The array beam-pattern can generally be designed to have a specified response. This can be done by properly setting the values of the multi-channel filters weights. Statistically optimal beamformers are designed based on the statistical properties of the de-

sired and interference signals. In general, they aim at enhancing the desired signals, while rejecting the interfering signals. Several criteria can be applied in the design of the beamformer, e.g., maximum signal to noise ratio (MSNR), minimum mean square error (MMSE), minimum variance distortionless response (MVDR) and LCMV. A summary of several design criteria can be found in [25, 26]. Cox et al. [27] introduced an improved adaptive beamformer that maintains a set of linear constraints as well as a quadratic inequality constraint.

In [28, 29] a multichannel Wiener filter (MWF) technique has been proposed that produces an MMSE estimate of the desired speech component in one of the microphone signals. The speech distortion weighted multichannel Wiener filter (SDW-MWF) [30][31], generalizes both MMSE and MVDR criteria. By identifying the two terms of error as distortion and residual noise, and weighting the residual noise component in the MMSE minimization by a factor μ, it is possible to control the tradeoff between the two error sources. By setting $\mu = 1$ or $\mu = 0$, the MWF and the MVDR are obtained as special cases of the SDW-MWF, respectively. Doclo et al. [32] show that the SDW-MWF is equivalent to the MVDR followed by a single channel SDW-MWF post-filter.

In an MVDR beamformer [33, 34], the power of the output signal is minimized under the constraint that signals arriving from the assumed direction of the desired speech source are processed without distortion. A widely studied adaptive implementation of this beamformer is the GSC [35]. The standard GSC consists of a spatial pre-processor, i.e. a fixed beamformer (FBF), also known as the quiescent filter, and a BM, combined with a multichannel adaptive noise canceler (ANC). The FBF provides a spatial focus on the speech source, creating a so-called speech reference; the BM steers nulls in the direction of the speech source, creating so-called noise references; and the multichannel ANC eliminates the noise components in the speech reference that are correlated with the noise references. Several researchers (e.g. Er and Cantoni [36]) have proposed modifications to the MVDR for dealing with multiple linear constraints, denoted LCMV. Their work was motivated by the desire to obtain further control of the array/beamformer beam-pattern, beyond that of a steer-direction gain constraint. Hence, the LCMV can be applied for constructing a beam-pattern satisfying certain constraints for a set of directions, while minimizing the array response in other directions. Breed and Strauss [37] proved that the LCMV has an equivalent GSC structure, which decouples the constraining and the minimization operations. The GSC structure was reformulated in the frequency domain, and extended to deal with the more complicated general ATFs case by Affes and Grenier [38] and later by Gannot et al. [39]. The latter

frequency-domain version, which takes into account the reverberant nature of the enclosure, was nicknamed the transfer function generalized sidelobe canceler (TF-GSC).

Several beamforming algorithms based on subspace methods were developed. Ephraim and Van Trees [40] considered the single microphone scenario. The eigenvalue decomposition (EVD) of the noisy speech correlation matrix is used to determine the signal and noise subspaces. Each of the eigenvalues of the signal subspaces is then processed to obtain the minimum distorted speech signal under a permissible level of residual noise at the output. Hu and Loizou [41] extended this method to deal with the colored noise case by using the generalized eigenvalue decomposition (GEVD) rather than the EVD as in the white noise case. Gazor et al. [42] proposed to use a beamformer based on the MVDR criterion and implemented as a GSC to enhance a narrowband signal contaminated by additive noise and received by multiple sensors. Under the assumption that the DOA entirely determines the transfer function relating the source and the microphones, it was shown that determining the signal subspace suffices for the construction of the algorithm. An efficient DOA tracking system, based on the projection approximation subspace tracking deflation (PASTd) algorithm [43] was derived. An extension to the wide-band case was presented by the same authors [44]. However the requirement for a delay-only impulse response is still not relaxed. Affes and Grenier [38] applied the PASTd algorithm to enhance speech signal contaminated by spatially white noise, where arbitrary ATFs relate the speaker and the microphone array. The algorithm was proved to be efficient in a simplified trading-room scenario, where the DRR is relatively high and the reverberation time relatively low. Doclo and Moonen [28] extended the structure to deal with the more complicated colored noise case by using the generalized singular value decomposition (GSVD) of the received data matrix. Warsitz et al. [45] proposed to replace the BM in [39]. They used a new BM based on the GEVD of the received microphone data, providing an indirect estimation of the ATFs relating the desired speaker and the microphones.

Affes et al. [46] extended the structure presented in [42] to deal with the multi-source case. Asano et al. [47] addressed the problem of enhancing multiple speech sources in a non-reverberant environment. The multiple signal classification (MUSIC) method, proposed by Schmidt [48], is utilized to estimate the number of sources and their respective steering vectors. The noise components are reduced by manipulating the generalized eigenvalues of the data matrix. Based on the subspace estimator, an LCMV beamformer is constructed. The LCMV constraints set consists of two subsets: one for maintaining the desired sources and the

second for mitigating the interference sources. Benesty et al. [49] also addressed beamforming structures for multiple input signals. In their contribution, derived in the time-domain, the microphone array is treated as a MIMO system. In their experimental study, it is assumed that the filters relating the sources and the microphones are a priori known, or alternatively, that the sources are not active simultaneously. Reuven et al. [50] dealt with the scenario in which one desired source and one competing speech source coexist in noisy and reverberant environment. The resulting algorithm, denoted dual source transfer function generalized sidelobe canceler (DTF-GSC) is tailored to the specific problem of two sources and cannot be easily generalized to the multiple desired and interference sources. Markovich-Golan et al. [51] considered the multiple speakers case and proposed to use an LCMV-BF in a GSC-form. The multichannel eigen-spaces BF for multiple sources in a reverberant scenario is briefly presented in Sec. 2.1. This LCMV beamformer satisfies two sets of linear constraints. One set is dedicated to maintaining the desired signals, while the other set is chosen to mitigate both the stationary and non-stationary interferences. The speakers are assumed static and the noise statistics is assumed slowly time-varying. Unlike classical beamformers, which approximate the room impulse response (RIR)s as delay-only filters, we take into account the entire RIR [or its respective ATF]. The LCMV beamformer is then reformulated in a GSC structure, consisting of a FBF, BM and ANC. We show that two sets of basis vectors, one spanning the desired speakers subspace and a second spanning the interference subspace suffice for constructing the beamformer. The basis estimations are performed by collecting eigenvectors, calculated in segments where desired or competing speakers are exclusively active. The rank of the two basis is then reduced by the applying the orthogonal triangular decomposition (QRD). This procedure relaxes the common requirement for non-overlapping activity periods of the interference sources.

Whereas satisfactory results are obtained with multiple static speakers, moving speakers enhancement remains a cumbersome and unsolved task. In chapter 2, we present an LCMV-BF which we developed in previous work. This BF serves as a starting point for a subspace tracking algorithm for dynamic scenarios. Subsequently, a novel beamforming criterion extending the SDW-MWF to multiple constraints is derived.

1.2 Distributed beamformers for Speech processing

The subject of distributed signal processing in sensor networks has been the focus of ongoing research in recent years. Distributed signal processing techniques aim at efficiently and accurately extracting information out of multiple sensors data, utilizing both temporal and spatial diversity of the network, and relying on nodes cooperation. Research fields include parameter estimation, tracking, localization and signal enhancement. Surveys on signal processing in WSNs can be found in [1, 2, 52, 53, 54, 55].

Although many contributions can be found on distributed signal processing algorithms, most of them focus on parameter estimation [56, 57, 58, 59] and localization and only few contributions consider signal enhancement applications, and even fewer deal with speech signals. In recent years, some contributions to the field of WASN have been introduced, circumventing the severe network constraints [60, 61, 62, 63, 64, 65, 66]. A trivial solution is obtained by utilizing only microphones local to the node without any communication link. However this solution fails to utilize the entire information from the network and hence is sub-optimal. Signals and parameters at a node, obtained by processing self-owned microphone signals, are denoted "local" to the node. Other signals and parameters shared by the nodes of the WASN are denoted "global". A common scheme for distributed signal processing algorithms in WASNs comprises the following steps. First, local processing of microphone signals results in intermediate signals or estimates at each node, requiring less communication-bandwidth. Second, the results of the first step are broadcast in the WASN. Finally, a global estimate or an enhanced signal is obtained by merging all intermediate signals or estimates. Since the data available at each node is incomplete, an iterative (or time-recursive) solution becomes a necessity.

An interesting application of distributed signal processing algorithms is binaural hearing aids. In regular hearing aids the right and left apparatuses operate independently, performing single/multi-channel noise reduction. We refer to this scheme as the *bilateral* hearing aids. On the contrary, binaural hearing aids share information between left and right laterals and obviously hold more potential for better speech understanding. A survey on binaural hearing aids can be found in [67].

A dual channel SDW-MWF has been proposed and analyzed as a *binaural* hearing scheme in [68, 69, 70, 71, 72]. Both SDW-MWFs beamformers are applied to microphone data from both sides and the scheme is denoted as the binaural multichannel Wiener filter (B-MWF).

The desired signals are chosen from a reference signal at each side. This scheme not only controls the tradeoff between noise reduction and signal distortion, it also controls the distortions of the interaural time difference (ITD) and the interaural level difference (ILD). Lower signal distortion results in lower distortion of the localization cues and higher noise levels.

Data sharing among two sides requires bi-directional transmission in both hearing aids as they are not connected by wires for convenience reasons. As advanced hearing aids are comprised of several microphones in each side, and as power limitations prevent transmission of all data, reduced bandwidth algorithms are called upon. Doclo et al. [62] proposed an iterative distributed SDW-MWF (DB-MWF) algorithm which reduces the required bandwidth of the binaural link to one channel per side regardless of the number of microphones. The authors proved that their scheme converges to the optimal B-MWF for the rank-1 spatial desired signal covariance matrix. Roy and Vetterli [73] analyzed the rate-distortion curve for the noise reduction problem in binaural hearing aids. They considered the performance bound for the case where the correlation between the lateral hearing aids is known, and for the more practical case where the correlation is unknown.

Bertrand and Moonen [63] considered the more general case of an N node WASN and P desired sources. They allowed each node to define individual desired signals by using different weighting of the spatial components of the speech. They proposed a DANSE-P algorithm which necessitates transmission of P channels from each node and proved the convergence of the algorithm to the global SDW-MWF BF. In complicated scenarios where multiple speakers exist and more control over the beampattern is desired, the LCMV-BF is a more suitable option. The linear constraints set can be designed to maintain undistorted desired speakers while mitigating competing speakers. Bertrand and Moonen [74] also proposed a distributed LCMV algorithm, denoted linearly constrained DANSE (LC-DANSE). They considered the case of P speakers and noise picked up by microphones of an N node WASN. Assuming that each node may define the set of desired speakers differently, they proposed that the constraints matrix will be common to all nodes, whereas the desired response will be node-specific. Their proposed algorithm constructs P node-specific constraints LCMV-BFs that require each node to transmit P audio channels. A total of $N \times P$ transmission channels (the output signals of all local BFs) are required. At each iteration, each node has to re-estimate two sets of basis vectors spanning the ATFs of the desired and the interfering speakers.

Special considerations for applying BSS in WASNs have been considered. Dmochowski

et al. [75] propose a method for alleviating the speakers overlap problem in teleconferencing. Their algorithm is based on frequency domain BSS, followed by a post-filter utilizing time segments in which only a subset of the speakers is active. Ono et al. [76] also apply frequency domain BSS algorithms in WASN. They use pairs of proximately located microphones in order to resolve permutation ambiguity. Assuming that there is no spatial aliasing in those pairs, the estimated phases between microphones are used for correctly combining the frequency components of each of the speakers.

As each node relies on its own clock source, synchronization offsets are inevitable. Elson and Kay [77] considered the time synchronization problem in WSNs. They defined several aspects of the problem, namely phase and frequency synchronization problems of the nodes' clock sources. They proposed an algorithm, denoted reference-broadcast synchronization (RBS), for synchronizing the clocks in the WSN. Ando and Ono [78] discussed the importance of synchronizing the time origin in WASNs. Ono et al. [79] considered such unsynchronized nodes (with a fixed delay uncertainty). They proposed a blind estimation method for these delays based on localizing the speakers and the microphones.

Wehr et al. [60] considered the synchronization problem in distributed beamforming for BSS. They proposed an algorithm for estimating the sampling rate offsets based on a modulated reference signal which is broadcast in the WASN. The estimated sampling rate offsets are then compensated for by resampling at the nodes. Pawig et al. [80] considered the problem of a sampling rate offset between the analog to digital converter (ADC) and the digital to analog converter (DAC) in a single channel echo cancelation system. They utilized the reference data to estimate the sampling rate offset, and proposed a combined time-recursive algorithm for tracking both the RIR and the sampling rate offset. Subsequently, the reference signal is resampled using the Lagrange polynomials interpolation method [81]. Liu [82] examined the performance degradation of BSS due to sampling rate offsets. Applying an ICA algorithm for the separation, he found a relation between the correlation of sources' energies (at the output of the BSS) and sampling rate offset. Based on the latter relation, an estimation procedure for the sampling rate offset was derived.

Liu and Chen et al. [83, 84] propose a method for localizing speakers and microphones based on energy in a distributed ad hoc WASN. Their method is less accurate than time difference of arrival (TDOA) methods, however it is more robust to synchronization errors between microphones.

Most distributed BFs adopt the MMSE criterion or perform data-independent DS-BF.

Still, only few consider the LCMV criterion and recursive algorithms. In chapter 3 a novel distributed GSC-BF for multiple speakers is derived. We also propose a synchronization method for estimating and compensating for sampling rate offsets.

1.3 Beamforming Properties in Random Sensor Networks

In classical beamforming applications the sensor array constellation is pre-determined to meet the required performance of the scenario at hand. Important attributes of array design include sensors spacing, directivity pattern, and of course number of sensors. Improper array design might lead to performance degradation due to small spatial diversity or grating lobes. The design parameters as well as the exact sensor locations are usually unknown to the signal processing algorithms. Van Veen and Buckley [25] discussed the beamformer design, and analyzeed its properties (such as directivity, gain and sidelobe level) with relation to the above physical parameters. In many distributed sensor network scenarios, pre-determining the array design is impossible. Nodes containing sensors are randomly distributed in many scenarios (for example scattered from an airplane), limited energy results in nodes shutdown, and in some cases nodes can move (for example in the sea and when attached to animals). These scenarios lead to uncontrollable layout of the sensor array. It is therefore important to replace classical array design with a different framework suitable to distributed sensor network in order to assure required performance. Treating sensor locations as random variables and assuming a dynamic network connectivity enable the analysis of beamforming performance in a statistic framework.

Lo [6] proposed to adopt a statistical model for the sensors positions and to explore the incurred distribution of beampattern properties. He treated the beamformer properties as functions of the random sensors locations and modeled their statistics. He examined the properties of beam-width, sidelobe level and gain. Yao et al. [85] proposed a blind beamformer, with unknown sensors positions based on maximizing the output power while constraining the norm of the weights. Ochiai et al. [4] extended the one dimensional analysis, and considered the scenario of distributed clusters of nodes (a cluster is a group of nodes closely located). They dealt with a closely related problem of steering the beampattern of a certain cluster towards another cluster. They assume that the antennas in each cluster are uniformly

distributed on a disk in a two dimensional plane, and analyze the properties of the beampattern. Kerby and Bernhard [86] considered a compound two dimensional array comprising of rotated replications of a smaller sub-array consisting of uniformly distributed antennas, arranged in a two dimensional matrix. They analyzed the statistics of the sidelobe level and the wideband behavior of the array. Ahmad and Vorobyov [87] analyzed the characteristics of a three dimensional cloud of Normally distributed sensors. All past analysis considered a simple DS beamformer in a non-reflective environment, where the desired source is in the far field regime. Kodrasi et al. [88] showed that the array performance depends significantly on the microphone locations and compared various heuristic array design optimization methods. They considered the superdirective BF, which maximizes the directivity index.

A performance analysis for data-dependent BFs with randomly located microphones in a reverberant environment has not been examined yet. In chapter 4 we analyze the performance of the SDW-MWF BF, utilizing randomly located microphones, in a reverberant environment.

1.4 Computational complexity of distributed beamforming

WASN applications introduce conflicting requirements for computational power, battery life, and price. On the one hand, vast sensor networks with numerous nodes and sensors require significantly increased computations. Furthermore, the dynamics of the network and the environment necessitates updating the BF frequently. The computational burden is emphasized in wideband signal applications such as speech processing in reverberant environments. Dealing with such long RIRs requires calculating BFs with respectively long impulse responses and involves many computations. On the other hand, strict constraints on power consumption and on node price limit the complexity of algorithms and the amount of data shared by nodes via wireless communication. Complex algorithms require stronger processing units resulting in more expensive nodes, and might prevent deployment of large quantities. Straightforward, centralized beamforming algorithms lack scalability, and do not meet these constraints. Hence, developing distributed algorithms to alleviate these contradicting requirements is called upon.

As mentioned earlier, constructing a BF utilizing all sensor data requires large bandwidth. The contribution of each of the nodes to the noise reduction task is not equal. Given

a bandwidth limitation, a subset of the nodes could be chosen to maximize the noise reduction. Bertrand and Moonen [89] proposed an efficient method for updating the MWF-BF corresponding to removal or addition of sensors. They derived equations for efficiently recalculating the MWF-BF based on the previous BF by applying the block-matrix inversion formula [90].

Next, we examine the structure of the commonly used GSC-BF and identify the main procedure contributing to its complexity. Assuming that M microphones are used, and K constraints are defined, the complexity per frame and frequency bin of the GSC in the short time Fourier transform (STFT) domain is attributed to: 1) M complex multiplications at the FBF; 2) the complexity of applying the BM; 3) $M-K$ complex multiplications at the noise canceler (NC). Note that the BM is the main contributor to the complexity, yet there is no unique design for it. For the single constraint scenario, Gannot et al. [39] proposed a sparse $M \times (M-1)$ BM which requires only $M-1$ complex multiplications per frequency. Herbordt and Kellerman [91] provided an efficient adaptive BM utilizing also only $M-1$ complex multiplications. A commonly used $M \times (M-K)$ BM is comprised of the basis vectors spanning the null-subspace of the constraints columns-space. The basis vectors can be obtained by applying the singular value decomposition (SVD) to the constraints matrix. Applying the SVD based BM involves $M \times (M-K)$ complex multiplications. Markovich-Golan et al. [51] used an $M \times M$ projection matrix to the null-subspace of the constraints matrix as the BM. The latter scheme requires M^2 complex multiplications. Reuven et al. [50] and Krueger et al. [92] proposed a 2-stage projection procedure for designing a BM for the case of a single desired speaker and a single interfering speaker. The resulting BM is an $M \times M$ matrix, and its application requires $2 \times M$ and M^2 complex multiplications in [50] and [92], respectively. Clearly, the complexity of the GSC is mainly dominated by the BM (in the general K constraints case). Tseng and Griffiths [93] proposed to construct the BM by recursively projecting the received signals to the null-space of the constraints, one by one. The resulting procedure requires MK complex multiplications (provided that $K < \frac{M}{2}$).

In chapter 5 efficient methods for implementing and updating distributed BFs are developed. We consider changes in the available microphones and in the sources activity. A substantial saving in computation is obtained by making use of previously calculated BFs. Additionally, an efficient implementation of a BM is proposed.

1.5 Dissertation structure

The structure of the dissertation follows next. We briefly introduce the chapter topics and their sections. We give references to our publications on each of the subjects.

In Chapter 2 we survey our previous work on beamforming criteria for speech processing which is a basis for the current work. The multichannel eigen-spaces BF for multiple sources in a reverberant scenario is briefly presented in Sec. 2.1. This beamforming algorithm is based on [51] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals", *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1071-1086, Aug. 2009). This method constitutes the basic building block of the algorithms presented in this dissertation.

In Sec. 2.2, an extension to the SDW-MWF for the case of multiple sources is derived. We identify the various error components at the output of the BF as residual noise and signal distortion. For each source, the distortion is defined as the variance of the error between the desired and actual responses. We propose to apply individual weights to each of the distortion components. We prove that the LCMV-BF is a special case of the proposed beamformer, denoted as multiple speech distortions weighted multichannel Wiener filter (MSDW-MWF). Although the derived algorithm is centralized, its flexibility in treating multiple sources makes it a good candidate for implementation as a distributed BF. This section is based on [94] (S. Markovich-Golan, S. Gannot, and I. Cohen, "A weighted multichannel Wiener filter for multiple sources scenarios" in *the 27th convention of the Israeli Chapter of IEEE*, Eilat, Israel, Nov. 2012).

In practice, the ATFs of the sources vary over time in WSN, corresponding to their movements, or to dynamics of the reflectors in the enclosure. In Sec. 2.3, we propose an extension to the LCMV-BF in [51] for dynamic scenarios. The proposed BF adopts the PASTd algorithm for tracking non-static scenarios, presented in [51], in which multiple speakers coexist in a reverberant environment. The proposed algorithm is capable of extracting a desired conversation out of many conversations in time-varying and reverberant scenarios, where the expected DRR can be low. This section is based on [95] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Subspace tracking of multiple sources and its application to speakers extraction", *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 201–204, Mar. 2010).

In Chapter 3, we develop distributed beamforming algorithms for various scenarios and applications. In Sec. 3.1 a novel reduced-bandwidth *iterative* binaural MVDR beamformer is proposed. The proposed method reduces the bandwidth requirement for communication between hearing aids to a single channel, regardless of the number of microphones. The algorithm is proven to converge to the optimal binaural MVDR in the case of a rank-1 desired source correlation matrix. This section is based on [64] (S. Markovich-Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer", in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010). In the following contributions, time-recursive rather than iterative distributed BFs are developed which operate on a batch of the received data. The time-recursion procedure is inherently more suitable for dynamic scenarios.

Sec. 3.2 is dedicated for presenting a distributed beamforming algorithm for the case of multiple speakers scenario. We consider the case where the nodes *agree* on the classification of desired and competing speakers and share a common constraints set as well as desired responses. A distributed time-recursive version of the centralized GSC, denoted distributed GSC (DGSC), is proposed. We prove that the proposed algorithm converges to the centralized GSC. The proposed algorithm requires the transmission of only $N + P$ audio channels. In static scenarios, the relative transfer functions (RTFs) of the sources need to be estimated only once at the initialization stage. The estimation procedure of the RTFs may require non-overlapping activity patterns of the speakers. This section is based on [96] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks", *IEEE Trans. Audio, Speech and Language Processing*, 2012).

In Sec. 3.3, a distributed BF for the special case of a single desired speaker is developed. The efficient GSC-form implementation of the MVDR, rather than its closed-form, is considered. The GSC-form relaxes the requirement of the LC-DANSE algorithm to re-estimate the speech and noise spectra at each iteration. Both iterative and time recursive procedures are derived. The proposed algorithm, denoted distributed single-constraint generalized sidelobe canceler (DS-GSC), is based on a two-stage GSC. In the first stage, N local GSC-BFs are applied only to the local microphones at each node, yielding N signals which are broadcast in the WASN (instead of $N + 1$ in the DGSC for this case). The second stage, comprises a global GSC BF which processes the N output signals of the first stage. A replica of the global filtering stage is applied simultaneously and independently in all the nodes of the WASN.

The main advantages of the proposed scheme are its ability to adapt in speech-absent segments, and that it relaxes the requirement of closed-form MVDR algorithms to estimate the speech spectrum repeatedly in static environments. We prove that the proposed algorithm converges to the centralized TF-GSC. This section is based on [97] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed GSC beamforming using the relative transfer function", in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug. 2012, pp. 1274–1278).

The proposed distributed algorithms thus far assumed that all microphone signals are sampled at an accurate and equal sampling rate at all nodes. In Sec. 3.4, relax this requirement and a blind procedure for estimating the sampling frequency offsets and compensating for them is derived. The procedure is applicable to scenarios with slowly time-varying noise signals during speech-absent time segments. The proposed procedure is based on the phase drift of the coherence between two signals sampled at different sampling rates. Resampling the signals with Lagrange polynomials interpolation method compensates for the sampling rate offsets. This section is based on [98] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming", in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aachen, Germany, Sep. 2012).

In Chapter 4 the topic of statistical BFs is considered. The performance of data-dependent BFs with randomly located microphone is analyzed in two important scenarios. Sec. 4.1, is dedicated to derivation of a novel statistical model for performance analysis of the MWF beamformer. We consider the scenario of one desired source and one interfering source arriving from the far-field and impinging on a uniformly distributed linear array, in a non-reverberant environment. A theoretical model for the MMSE is developed and verified by simulations. The applicability of the proposed statistical model for speech signals is discussed. This section is based on [99] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Performance analysis of a randomly spaced wireless microphone array", in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011, pp. 121–124).

In Sec. 4.2, we further consider applying the SDW-MWF (which generalized the MVDR) to enhance a desired source propagating in a reverberant enclosure where the microphones are randomly located with a uniform distribution. Two noise fields are considered, namely, multiple coherent interference signals and a diffuse noise. Utilizing the statistics of the ATF, we derive a statistical model for two important criteria of the BF: the SIR, and the white noise gain. Moreover, we propose reliability functions, which determine the probability of the

SIR and sensitivity to exceed a desired level. This section is based on [100] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure" submitted for publication in *IEEE Trans. Audio, Speech and Language Processing*, Aug. 2012).

The computational complexity of beamforming algorithms can be quite high, as presented in previous sections. Due to the strict limitations in WASN it is desired to reduce this complexity. In Chapter 5 we propose procedures for reducing the computational complexity of applying beamforming in WASNs. In Sec. 5.1, a novel systematic scheme for constructing a multiple constraints sparse BM is presented. The sparsity of the proposed BM substantially reduces the complexity to $K \times (M - K)$ complex multiplications, where K is the number of constraints. A theoretical analysis of the signal leakage and of the blocking ability of the proposed sparse BM and of the eigen-space BM is derived. It is proven analytically, and tested for narrowband signals and for speech signals, that the blocking abilities of the sparse and of the eigen-space BMs are equivalent. This section is based on [101] (S. Markovich-Golan, S. Gannot, and I. Cohen, "A sparse blocking matrix for multiple constraints GSC beamformer", *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2012).

In Sec. 5.2, we consider sub-optimal LCMV beamformers utilizing only a subset of the available sensors for signal enhancement applications. Multiple desired and interfering sources scenarios in multi-path environments are considered. We assume that an *oracle* entity determines the group of sensors participating in the spatial filtering, denoted as the active sensors. The oracle is also responsible for updating the constraints set according to either sensors or sources activity or dynamics. Any update of the active sensors or of the constraints set necessitates re-calculation of the beamformer and increases the power consumption. As power consumption is one of the most valuable resources in sensor networks, it is important to derive efficient update schemes. In this section we derive procedures for adding or removing either an active sensor or a constraint from an existing LCMV beamformer. Closed-form as well as GSC-form implementations are derived. These procedures use the previous beamformer to save calculations in the updating process. We analyze the computational burden of the proposed procedures and show that it is much lower than the computational burden of the straightforward calculation of their corresponding beamformers. This section is based on [102] (S. Markovich-Golan, S. Gannot, and I. Cohen, "Low-complexity addition or removal of sensors/constraints in LCMV beamformers", *IEEE Transactions on Signal Processing*, vol.

60, no. 3, pp. 1205–1214, Mar. 2012).

1.6 List of publications

The list of publications of the author included in this dissertation are enlisted next.

Journal publications

1. S. Markovich-Golan, S. Gannot, and I. Cohen, "Low-complexity addition or removal of sensors/constraints in LCMV beamformers", *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1205–1214, Mar. 2012.

2. S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks", *IEEE Trans. Audio, Speech and Language Processing*, 2012.

3. S. Markovich-Golan, S. Gannot, and I. Cohen, "Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure" accepted for publication in *IEEE Trans. Audio, Speech and Language Processing*, Mar. 2013.

Book chapters

1. S. Markovich-Golan, Sharon Gannot and Israel Cohen, *Extraction of Desired Speech Signals in Multiple-Speakers Reverberant Noisy Environments*, to appear in Speech Processing in Modern Communication: Challenges and Perspectives, 2009. Chapter 10, pp. 255–280

Peer reviewed conference publications

1. S. Markovich-Golan, S. Gannot, and I. Cohen, "Subspace tracking of multiple sources and its application to speakers extraction", *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 201–204, Mar. 2010.

2. S. Markovich-Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer", in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010 [**Best student paper award**].

3. S. Markovich-Golan, S. Gannot, and I. Cohen, "Performance analysis of a randomly spaced wireless microphone array", in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011, pp. 121–124.

4. S. Markovich-Golan, S. Gannot, and I. Cohen, "A sparse blocking matrix for multiple constraints GSC beamformer", *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2012.

5. S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed GSC beamforming using the relative transfer function", in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug. 2012, pp. 1274–1278.

6. S. Markovich-Golan, S. Gannot, and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming", in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aachen, Germany, Sep. 2012 [**Short-list for the best student paper award**].

7. S. Markovich-Golan, S. Gannot, and I. Cohen, "A weighted multichannel Wiener filter for multiple sources scenarios", in *the 27th convention of the Israeli Chapter of IEEE*, Eilat, Israel, Nov. 2012 [**Best student paper award**].

Chapter 2

Beamforming algorithms for speech processing

The problem of extracting several desired speech signals, contaminated by non-stationary and stationary interfering signals is considered. We survey a BF for multiple speakers scenario which was developed in our previous work [51]. This BF serves a common basis for other contributions derived in this dissertation. The BF is based on the LCMV criterion and suggests to associate a constraint with a desired response to each of the speakers (both desired and intefering). A novel estimation method relaxes the requirement for non-concurrent speakers' activity patterns in RTFs estimation. We proposed a procedure of estimating a basis which spans the RTFs, for a more general activity patterns. This previous work is briefly reviewed in Sec. 2.1.

In Sec. 2.2, an extension of the SDW-MWF BF to the case of multiple speakers is proposed. We derive a procedure that enables control of the distortion of each of the sources at the output of the BF. It is further shown that the proposed design generalizes the LCMV-BF.

Both BF design assume static speakers. In Sec. 2.3, an algorithm for tracking the basis of the desired speakers' ATF, and the basis of the interfering speakers' ATFs is presented. The algorithm is incorporated in the LCMV-BF resulting in an adaptive version.

2.1 Eigen-spaces LCMV-BF

In this section, we review an LCMV beamformer designed for extracting desired speech signals from multi-microphone measurements, first presented in [51]. The BF satisfies two sets of linear constraints. One set is dedicated to maintaining the desired signals, while the other set

is chosen to mitigate both the stationary and non-stationary interferences. Unlike classical BFs, which approximate the RIRs as delay-only filters, we take into account the entire RIR [or its respective ATF]. The LCMV-BF is then reformulated in a GSC structure, consisting of a FBF, BM and ANC. For spatially-white noise field, the BF reduces to a FBF, satisfying the constraint sets, without power minimization. The application of the ANC contributes to interference reduction, only when the constraint sets are not completely satisfied. We show that two basis, one spanning the desired speech sources' ATFs and second spanning the interfering sources' ATFs, suffice for constructing the BF. A basis for the interference subspace is estimated by collecting eigenvectors, calculated in time segments where non-stationary interfering sources are active (with arbitrary activity) and the desired sources are inactive. The rank of the basis is then reduced by the application the QRD. Similarly, the basis of the desired ATFs is estimated by collecting generalized eigenvectors during time-segments in which only desired speakers are active (with arbitrary activity). By applying the GEVD procedure to the power spectral density (PSD) matrices the noise covariance matrix is whitened and we obtain a basis which spans the ATFs of the desired sources. This procedure relaxes the common requirement for non-overlapping activity periods of the various sources. This BF is denoted as the eigen-spaces LCMV-BF due to its estimation procedure of the constraints matrix.

This section is organized as follows. In Sec. 2.1.1 we introduce the mathematical formulation, describing the problem and in Sec. 2.1.2 we introduce the eigen-spaces LCMV-BF.

2.1.1 Problem Formulation

Consider a scenario with multiple simultaneous conversations. N_d desired speakers $s_{d,1}(n),\ldots,s_{d,N_d}(n)$ and N_i competing speakers $s_{i,1}(n),\ldots,s_{i,N_i}(n)$ are received in a noisy and reverberant environment by an M microphones array. The received signals are formulated in a vector notation in the STFT domain:

$$\boldsymbol{z}(\ell,k) \triangleq \boldsymbol{H}_d(\ell,k)\boldsymbol{s}_d(\ell,k) + \boldsymbol{H}_i(\ell,k)\boldsymbol{s}_i(\ell,k) + \boldsymbol{v}(\ell,k) \tag{2.1}$$

where $\boldsymbol{s}_d(\ell,k) \triangleq \left[\begin{array}{ccc} s_{d,1}(\ell,k) & \cdots & s_{d,N_d}(\ell,k)\end{array}\right]^T$ and $\boldsymbol{s}_i(\ell,k) \triangleq \left[\begin{array}{ccc} s_{i,1}(\ell,k) & \cdots & s_{i,N_i}(\ell,k)\end{array}\right]^T$ are vectors comprising the desired and interfering speakers, respectively. $\boldsymbol{v}(\ell,k)$ denotes the noise component, consisting of a mixture of spatially white, directional and diffuse noise sources. We further assume that the statistics of this noise component is slowly

time-varying. k denotes the frequency-bin index and ℓ the frame index. $\boldsymbol{H}_d(\ell,k) \triangleq \begin{bmatrix} \boldsymbol{h}_{d,1}(\ell,k) & \cdots & \boldsymbol{h}_{d,N_d}(\ell,k) \end{bmatrix}$ and $\boldsymbol{H}_i(\ell,k) \triangleq \begin{bmatrix} \boldsymbol{h}_{i,1}(\ell,k) & \cdots & \boldsymbol{h}_{i,N_i}(\ell,k) \end{bmatrix}$ are $M \times N_d$ and $M \times N_i$ matrices, respectively, comprised of the time-varying ATFs relating desired and competing speakers with the microphone array. Henceforth, we assume that the speakers and the enclosure are static, and omit the index ℓ from the ATFs and the second order moments. Furthermore, the formulas are derived for a specific frequency-bin index k, and are similar for all other frequencies, hence the frequency-bin index is also omitted for brevity.

Assuming that the sources and the various interference signals are uncorrelated, the spatial correlation matrix of the received signals is formulated as:

$$\boldsymbol{\Phi}_{zz} \triangleq \boldsymbol{H}_d \boldsymbol{\Lambda}_d \boldsymbol{H}_d^H + \boldsymbol{H}_i \boldsymbol{\Lambda}_i \boldsymbol{H}_i^H + \boldsymbol{\Phi}_{vv} \tag{2.2}$$

where $\boldsymbol{\Lambda}_d \triangleq \text{diag}\left(\begin{bmatrix} \sigma_{d,1}^2 & \dots & \sigma_{d,N_d}^2 \end{bmatrix}\right)$ and $\boldsymbol{\Lambda}_i \triangleq \text{diag}\left(\begin{bmatrix} \sigma_{i,1}^2 & \dots & \sigma_{i,N_i}^2 \end{bmatrix}\right)$ are diagonal matrices with the spectral variances of the desired and interfering speakers on their main diagonal, respectively, and $\boldsymbol{\Phi}_{vv}$ is the noise spatial correlation matrix.

Denote the BF by $\boldsymbol{w}$, its output is defined as:

$$y(\ell) \triangleq \boldsymbol{w}^H \boldsymbol{z}(\ell). \tag{2.3}$$

2.1.2 BF structure

The eigen-spaces LCMV-BF is designed to reproduce the desired speakers components as received by a reference microphone, while canceling the competing speakers components and minimizing the overall residual noise power. Define the LCMV criterion with multiple constraints on both the desired and competing speakers:

$$\boldsymbol{w} = \underset{\boldsymbol{w}}{\text{argmin}}\{\boldsymbol{w}^H \boldsymbol{\Phi}_{vv} \boldsymbol{w}\};\ \text{s.t.}\ \boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{g}. \tag{2.4}$$

Define the linear constraints set:

$$\boldsymbol{C} \triangleq \begin{bmatrix} \boldsymbol{H}_d & \boldsymbol{H}_i \end{bmatrix} \tag{2.5a}$$

$$\boldsymbol{g} \triangleq \begin{bmatrix} \boldsymbol{1}_{1\times N_d} & \boldsymbol{0}_{1\times N_i} \end{bmatrix}^T \tag{2.5b}$$

$$\boldsymbol{C}^H \boldsymbol{w} = \boldsymbol{g} \tag{2.5c}$$

where $\boldsymbol{g}$ is the desired response vector and $\boldsymbol{C}$ is the constrains matrix. The closed-form solution for the minimization in (2.4) is given by:

$$\boldsymbol{w} = \boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{C}\left(\boldsymbol{C}^{H}\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{C}\right)^{-1}\boldsymbol{g}. \tag{2.6}$$

In practice, the ATFs are not available, and estimating them is a cumbersome task. Extending the work of Gannot et al. [39] to the multiple speakers scenario, we propose to construct the BF based on the RTFs, thus sacrificing dereverberation capability of the derived BF.

The RTF of a source, is defined as the ratio between the ATFs relating the source and the microphones and the ATF relating the source and a reference microphone (arbitrarily chosen here to be the microphone #1). Denote the RTF matrices of the desired and competing speakers as $\tilde{\boldsymbol{H}}_d$, $\tilde{\boldsymbol{H}}_i$ where:

$$\tilde{\boldsymbol{H}}_d \triangleq \left[\begin{array}{ccc} \tilde{\boldsymbol{h}}_{d,1} & \cdots & \tilde{\boldsymbol{h}}_{d,N_d} \end{array}\right] \tag{2.7a}$$

$$\tilde{\boldsymbol{H}}_i \triangleq \left[\begin{array}{ccc} \tilde{\boldsymbol{h}}_{i,1} & \cdots & \tilde{\boldsymbol{h}}_{i,N_i} \end{array}\right] \tag{2.7b}$$

where

$$\tilde{\boldsymbol{h}}_{d,p} \triangleq \frac{\boldsymbol{h}_{d,p}}{h_{d,p(1)}};\ p = 1,\ldots,N_d \tag{2.8a}$$

$$\tilde{\boldsymbol{h}}_{i,p} \triangleq \frac{\boldsymbol{h}_{i,p}}{h_{i,p(1)}};\ p = 1,\ldots,N_i. \tag{2.8b}$$

are the corresponding RTFs relating each of the sources and the microphone array.

Define the modified constraints set, based on the RTFs:

$$\tilde{\boldsymbol{C}} \triangleq \left[\begin{array}{cc} \tilde{\boldsymbol{H}}_d & \tilde{\boldsymbol{H}}_i \end{array}\right] \tag{2.9a}$$

$$\tilde{\boldsymbol{C}}^{H}\tilde{\boldsymbol{w}} = \boldsymbol{g}. \tag{2.9b}$$

Similarly to (2.6), the modified LCMV-BF equals:

$$\tilde{\boldsymbol{w}} = \boldsymbol{\Phi}_{vv}^{-1}\tilde{\boldsymbol{C}}\left(\tilde{\boldsymbol{C}}^{H}\boldsymbol{\Phi}_{vv}^{-1}\tilde{\boldsymbol{C}}\right)^{-1}\boldsymbol{g}. \tag{2.10}$$

Practical estimation procedures for the RTFs exist [51, 39], however they require time-segments in which each of the speakers is exclusively active. The eigen-spaces based LCMV-

BF relaxes the latter requirement, and replaces it with a requirement for time-segments in which the desired group and the competing group of speakers are exclusively active.

We propose a procedure for estimating a basis of the desired speakers ATFs based on the following steps: 1) estimate the noise covariance matrix during noise-only time segments; 2) estimate the covariance matrices during time-segments in which only the desired speakers and noise are active; 3) calculate the GEVD of each of the covariance matrices, and collect the vectors which correspond to the desired speakers subspace; 4) reduce the collected basis over multiple time-segments by applying the QRD. Estimating a basis for the competing speakers ATFs subspace is based on similar steps, only replacing the GEVD with an EVD. These procedures are discussed in further detail in Sec. 2.3.

Denote by $\boldsymbol{Q}_d \triangleq \left[\begin{array}{ccc} \boldsymbol{q}_{d,1} & \dots & \boldsymbol{q}_{d,N_d} \end{array} \right]$ a basis spanning the desired speakers subspace and by $\boldsymbol{Q}_i \triangleq \left[\begin{array}{ccc} \boldsymbol{q}_{i,1} & \dots & \boldsymbol{q}_{i,N_i} \end{array} \right]$ a basis spanning the competing speakers subspace. It can be shown (see Sec. 2.3) that the eigen-spaces based LCMV-BF, defined as:

$$\dot{\boldsymbol{w}} = \boldsymbol{\Phi}_{vv}^{-1} \dot{\boldsymbol{C}} \left(\dot{\boldsymbol{C}}^H \boldsymbol{\Phi}_{vv}^{-1} \dot{\boldsymbol{C}} \right)^{-1} \dot{\boldsymbol{g}} \tag{2.11}$$

is equivalent to the RTF-based LCMV-BF (2.10), where the eigen-spaces based constraints set is:

$$\dot{\boldsymbol{C}} \triangleq \left[\begin{array}{cc} \boldsymbol{Q}_d & \boldsymbol{Q}_i \end{array} \right] \tag{2.12a}$$

$$\dot{\boldsymbol{g}} \triangleq \left[\begin{array}{cccc} \boldsymbol{q}_{d,1}(1) & \cdots & \boldsymbol{q}_{d,N_d}(1) & \boldsymbol{0}_{1\times N_i} \end{array} \right]^T \tag{2.12b}$$

$$\dot{\boldsymbol{C}}^H \dot{\boldsymbol{w}} = \dot{\boldsymbol{g}}. \tag{2.12c}$$

2.2 A weighted multichannel Wiener filter for multiple sources scenarios

In this section, the same scenario as in Sec. 2.1 is considered (P speakers received by an M microphone array in a reverberant enclosure). We extend the single source SDW-MWF to deal with multiple speakers. The mean squared error (MSE) is extended by introducing P weights, each controlling the distortion of one of the sources. Opposed to LCMV criterion, the proposed method allows for controlled deviation from the desired responses. Thus, the proposed criterion leads to a more flexible BF design. Two special cases of the proposed BF

are the SDW-MWF [30],[70] and the LCMV-BF. We provide a theoretical analysis for the performance of the proposed BF. Finally, we exemplify the ability of the proposed method to control the tradeoff between NR and distortion levels of various speakers in an experimental study.

2.2.1 Problem formulation

The problem is formulated in the STFT domain, where ℓ and k are time-frame and frequency bin indices, respectively. Consider a microphone array located in a reverberant enclosure. The signals received by the microphone array are categorized in two groups. The first group comprises sources for which a desired response is designated. A source belonging to this group is denoted signal of interest (SOI). The second group comprises interferences that we wish to mitigate. Consider P coherent SOIs, denoted $s_1(\ell,k),\ldots,s_P(\ell,k)$. Denote by $\boldsymbol{h}_p(\ell,k)$, for $p=1,\ldots,P$, the ATF relating the pth SOI and the microphone signals. The received microphone signals are given by:

$$\boldsymbol{z}(\ell,k) \triangleq \boldsymbol{H}(\ell,k)\,\boldsymbol{s}(\ell,k) + \boldsymbol{v}(\ell,k) \tag{2.13}$$

where $\boldsymbol{s}(\ell,k) \triangleq \begin{bmatrix} s_1(\ell,k) & \cdots & s_P(\ell,k) \end{bmatrix}^T$ is a vector comprising all the SOIs, $\boldsymbol{H}(\ell,k) \triangleq \begin{bmatrix} \boldsymbol{h}_1(\ell,k) & \cdots & \boldsymbol{h}_P(\ell,k) \end{bmatrix}$ is an $M \times P$ matrix of the ATFs relating the SOIs and the microphones and $\boldsymbol{v}(\ell,k)$ denotes the received interferences. Next, we define the covariance matrices of the SOIs and interfering signals as $\boldsymbol{\Phi}_{ss}(\ell,k) \triangleq \mathrm{E}\left\{\boldsymbol{s}(\ell,k)\,\boldsymbol{s}^H(\ell,k)\right\}$ and $\boldsymbol{\Phi}_{vv}(\ell,k) \triangleq \mathrm{E}\left\{\boldsymbol{v}(\ell,k)\,\boldsymbol{v}^H(\ell,k)\right\}$, respectively. For brevity, hereafter the frequency bin index k is omitted and all derivations are valid for all $k=1,\ldots,K$ frequency bins. Moreover, we omit the frame index from $\boldsymbol{H}$, $\boldsymbol{\Phi}_{ss}$ and $\boldsymbol{\Phi}_{vv}$. The covariance matrix of the received signals is given by:

$$\boldsymbol{\Phi}_{zz} \triangleq \boldsymbol{H}\boldsymbol{\Phi}_{ss}\boldsymbol{H}^H + \boldsymbol{\Phi}_{vv}. \tag{2.14}$$

The desired response vector is denoted by $\boldsymbol{g}$ and the desired signal at the output of the BF is defined as:

$$d(\ell) \triangleq \boldsymbol{g}^H\boldsymbol{s}(\ell). \tag{2.15}$$

The output of a BF $\boldsymbol{w}$ is denoted by:

$$y(\ell) = \boldsymbol{w}^H \boldsymbol{z}(\ell) \tag{2.16}$$

and the MSE between the desired signal (2.15) and the BF's output (2.16) is:

$$J_w \triangleq \mathrm{E}\left\{|d(\ell) - y(\ell)|^2\right\}. \tag{2.17}$$

In the following section, we present the proposed algorithm.

2.2.2 Multiple speech distortions weighted multichannel Wiener filter

Substituting (2.13),(2.15),(2.16) in (2.17) and noting that $\boldsymbol{s}(\ell)$ and $\boldsymbol{v}(\ell)$ are statistically independent signals yields:

$$J_w = \left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}\right)^H \boldsymbol{\Phi}_{ss} \left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}\right) + \boldsymbol{w}^H \boldsymbol{\Phi}_{vv} \boldsymbol{w}. \tag{2.18}$$

We denote the component $\left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}\right)^H \boldsymbol{\Phi}_{ss} \left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}\right)$ as the total distortion and the component $\boldsymbol{w}^H \boldsymbol{\Phi}_{vv} \boldsymbol{w}$ as the residual noise. The SDW-MWF criterion introduces the parameter μ which controls the tradeoff between the total distortion and the noise reduction:

$$\begin{aligned} J_{\text{SDW-MWF}} \triangleq \min_{\boldsymbol{w}'} & \left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}'\right)^H \boldsymbol{\Phi}_{ss} \left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}'\right) \\ & + \mu \left(\boldsymbol{w}'\right)^H \boldsymbol{\Phi}_{vv} \boldsymbol{w}'. \end{aligned} \tag{2.19}$$

In this section, we propose to utilize individual parameters, one for each source, for controlling the distortion of each of the sources separately. Explicitly, the proposed MSE criterion is given by extending (2.19):

$$\begin{aligned} J_{\text{MSDW-MWF}} \triangleq \min_{\boldsymbol{w}'} & \left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}'\right)^H \boldsymbol{\Lambda} \boldsymbol{\Phi}_{ss} \left(\boldsymbol{g} - \boldsymbol{H}^H \boldsymbol{w}'\right) \\ & + \left(\boldsymbol{w}'\right)^H \boldsymbol{\Phi}_{vv} \boldsymbol{w}' \end{aligned} \tag{2.20}$$

where $\boldsymbol{\Lambda} \triangleq \mathrm{diag}\{\lambda_1, .., \lambda_P\}$, a diagonal matrix with the parameters λ_p for $p = 1, .., P$ on its diagonal, and the BF which minimizes (2.20) is denoted $\boldsymbol{w}_{\text{MSDW-MWF}}$, the MSDW-MWF.

The closed-form solution of (2.20) is given by:

$$\boldsymbol{w} = \left(\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{\Phi}_{ss}\boldsymbol{H}^H + \boldsymbol{\Phi}_{vv}\right)^{-1}\boldsymbol{H}\boldsymbol{\Lambda}\boldsymbol{\Phi}_{ss}\boldsymbol{g}. \tag{2.21}$$

Note, that for a single desired speaker scenario the SDW-MWF can be obtained as special case of the MSDW-MWF by setting:

$$\boldsymbol{\Lambda} = \mu^{-1}\boldsymbol{I}_{P\times P} \tag{2.22}$$

where $\boldsymbol{I}_{P\times P}$ is a $P \times P$ identity matrix. In the following sections we analyze the distortion of the SOIs and the noise level at the output of the proposed BF. In Sec. 2.2.2.3 we show that the well-known LCMV-BF is also a special case of the MSDW-MWF.

2.2.2.1 Distortion analysis

Two distortion figures of merit are analyzed. The first is the total distortion, defined as:

$$\begin{aligned} D_T \triangleq & \mathrm{E}\left\{|d(\ell) - y(\ell)|^2\right\} \\ = & \left(\boldsymbol{g} - \boldsymbol{H}^H\boldsymbol{w}\right)^H \boldsymbol{\Phi}_{ss}\left(\boldsymbol{g} - \boldsymbol{H}^H\boldsymbol{w}\right) \\ = & \|\boldsymbol{\Phi}_{ss}^{1/2}\left(\boldsymbol{g} - \boldsymbol{H}^H\boldsymbol{w}\right)\|^2 \end{aligned} \tag{2.23}$$

where $\boldsymbol{\Phi}_{ss} = \left(\boldsymbol{\Phi}_{ss}^{1/2}\right)^H\boldsymbol{\Phi}_{ss}^{1/2}$ is the Cholesky decomposition. The second is the individual distortion of the pth source for $p = 1, .., P$:

$$\begin{aligned} D_p \triangleq & \mathrm{E}\left\{|g_p^* s_p(\ell) - \boldsymbol{w}^H\boldsymbol{h}_p s_p(\ell)|^2\right\} \\ = & |g_p - \boldsymbol{h}_p^H\boldsymbol{w}|^2\phi_{ss,p} \end{aligned} \tag{2.24}$$

where $\phi_{ss,p}$ is the variance of the pth source, and $\boldsymbol{\Phi}_{ss} \triangleq \mathrm{diag}\left\{\phi_{ss,1}, .., \phi_{ss,P}\right\}$. Note, that since $\boldsymbol{\Phi}_{ss}$ is diagonal, $\boldsymbol{\Phi}_{ss}^{1/2}$ is also diagonal and therefore $D_p = \left|\left(\boldsymbol{\Phi}_{ss}^{1/2}\left(\boldsymbol{g} - \boldsymbol{H}^H\boldsymbol{w}\right)\right)_p\right|^2$, where $(\bullet)_p$ denotes the pth element of a vector.

Let

$$\boldsymbol{H} = \boldsymbol{U}\boldsymbol{S}\boldsymbol{V}^H \tag{2.25}$$

be the SVD of $\boldsymbol{H}$. Substituting (2.25) in (2.21) yields:

$$\boldsymbol{w} = \boldsymbol{U}\left(\boldsymbol{S}\boldsymbol{V}^H\boldsymbol{\Lambda}\boldsymbol{\Phi}_{ss}\boldsymbol{V}\boldsymbol{S}^H + \boldsymbol{U}^H\boldsymbol{\Phi}_{vv}\boldsymbol{U}\right)^{-1}\boldsymbol{S}\boldsymbol{V}^H\boldsymbol{\Lambda}\boldsymbol{\Phi}_{ss}\boldsymbol{g}. \tag{2.26}$$

Note that $\boldsymbol{S}$ is an $M \times P$ matrix of the form:

$$\boldsymbol{S} \triangleq \begin{bmatrix} \boldsymbol{S}_1 \\ \boldsymbol{0}_{(M-P)\times P} \end{bmatrix} \tag{2.27}$$

where $\boldsymbol{S}_1$ is a $P \times P$ diagonal real matrix. Hence, the expression $\boldsymbol{S}\boldsymbol{V}^H\boldsymbol{\Lambda}\boldsymbol{\Phi}_{ss}\boldsymbol{V}\boldsymbol{S}^H$ in (2.26) equals:

$$\boldsymbol{S}\boldsymbol{V}^H\boldsymbol{\Lambda}\boldsymbol{\Phi}_{ss}\boldsymbol{V}\boldsymbol{S}^H = \left[\begin{array}{c|c} \boldsymbol{S}_1\boldsymbol{V}^H\boldsymbol{\Lambda}\boldsymbol{\Phi}_{ss}\boldsymbol{V}\boldsymbol{S}_1 & \boldsymbol{0}_{P\times(M-P)} \\ \hline \boldsymbol{0}_{(M-P)\times P} & \boldsymbol{0}_{(M-P)\times(M-P)} \end{array}\right]. \tag{2.28}$$

Let $\boldsymbol{U}_1$ be an $M \times P$ matrix comprising the first P columns of U which span the column-space of $\boldsymbol{H}$, and let $\boldsymbol{U}_0$ be an $M \times (M-P)$ matrix comprising of the last $M-P$ columns of $\boldsymbol{U}$ which span the null-space of $\boldsymbol{H}$. I.e.,

$$\boldsymbol{U} = \begin{bmatrix} \boldsymbol{U}_1 & \boldsymbol{U}_0 \end{bmatrix}. \tag{2.29}$$

By substituting (2.29) in the expression $\boldsymbol{U}^H\boldsymbol{\Phi}_{vv}\boldsymbol{U}$ in (2.26), we obtain the following block-matrix structure:

$$\boldsymbol{U}^H\boldsymbol{\Phi}_{vv}\boldsymbol{U} = \left[\begin{array}{c|c} \boldsymbol{\Gamma}_A & \boldsymbol{\Gamma}_B \\ \hline \boldsymbol{\Gamma}_B^H & \boldsymbol{\Gamma}_C \end{array}\right] \tag{2.30}$$

where we define:

$$\boldsymbol{\Gamma}_A \triangleq \boldsymbol{U}_1^H\boldsymbol{\Phi}_{vv}\boldsymbol{U}_1 \tag{2.31a}$$

$$\boldsymbol{\Gamma}_B \triangleq \boldsymbol{U}_1^H\boldsymbol{\Phi}_{vv}\boldsymbol{U}_0 \tag{2.31b}$$

$$\boldsymbol{\Gamma}_C \triangleq \boldsymbol{U}_0^H\boldsymbol{\Phi}_{vv}\boldsymbol{U}_0. \tag{2.31c}$$

Now, applying the block-matrix inversion formula to the sum of (2.28) and (2.30) and substituting in (2.26) yields the following simplified expression:

$$\boldsymbol{w} = \boldsymbol{\Psi}\left(\boldsymbol{I} + \boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}_{ss}^{-1}\boldsymbol{\Theta}\right)^{-1}\boldsymbol{g} \tag{2.32}$$

where

$$\boldsymbol{\Psi} \triangleq \left(\boldsymbol{U}_1 - \boldsymbol{U}_0\boldsymbol{\Gamma}_C^{-1}\boldsymbol{\Gamma}_B^H\right)\boldsymbol{S}_1^{-1}\boldsymbol{V}^H \tag{2.33a}$$

$$\boldsymbol{\Theta} \triangleq \boldsymbol{V}\boldsymbol{S}_1^{-1}\left(\boldsymbol{\Gamma}_A - \boldsymbol{\Gamma}_B\boldsymbol{\Gamma}_C^{-1}\boldsymbol{\Gamma}_B^H\right)\boldsymbol{S}_1^{-1}\boldsymbol{V}^H. \tag{2.33b}$$

A more simplified expression can be obtained for cases in which low distortion is required. In these cases $\|\boldsymbol{\Lambda}\| \gg 1$, hence, we can assume that $\|\boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}_{ss}^{-1}\boldsymbol{\Theta}\| \ll 1$, and replace $\left(\boldsymbol{I} + \boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}_{ss}^{-1}\boldsymbol{\Theta}\right)^{-1}$ in (2.32) with its first order Taylor series approximation:

$$\boldsymbol{w} \approx \boldsymbol{\Psi}\left(\boldsymbol{I} - \boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}_{ss}^{-1}\boldsymbol{\Theta}\right)\boldsymbol{g}. \tag{2.34}$$

Finally, the total distortion is obtained by substituting (2.32) in (2.23):

$$D_T = \|\boldsymbol{\Phi}_{ss}^{1/2}\left(\boldsymbol{I} - \left(\boldsymbol{I} + \boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}_{ss}^{-1}\boldsymbol{\Theta}\right)^{-1}\right)\boldsymbol{g}\|^2. \tag{2.35}$$

Applying to (2.35) a similar approximation as in (2.34) yields an approximated expression for low distortion:

$$D_T \approx \|\boldsymbol{\Lambda}^{-1}\boldsymbol{\Phi}_{ss}^{-1/2}\boldsymbol{\Theta}\boldsymbol{g}\|^2. \tag{2.36}$$

Considering (2.36) and the relation between D_p and D_T, the following approximation holds:

$$D_p \approx \frac{|\boldsymbol{\theta}_p^H\boldsymbol{g}|^2}{\lambda_p^2\phi_{ss,p}} \tag{2.37}$$

where $\boldsymbol{\theta}_p$ is the pth column of the matrix $\boldsymbol{\Theta}$. Next, we define the various sources distortion measures. Define d_p as the distortion level of the pth source normalized by its power:

$$d_p \triangleq \frac{D_p}{\phi_{ss,p}}. \tag{2.38}$$

Define the set of desired distortion levels as $\dot{d}_p$ for $p = 1, \ldots, P$. Given such a set of desired distortion levels, a BF which satisfies them and minimizes the noise level can be obtained by using the proposed MSDW-MWF (2.21) with a proper $\boldsymbol{\Lambda}$ matrix whose diagonal elements

are given by:

$$\lambda_p = \frac{|\boldsymbol{\theta}_p^H \boldsymbol{g}|}{\sqrt{\dot{d}_p \phi_{ss,p}}};\ p = 1, ..P. \tag{2.39}$$

2.2.2.2 Noise analysis

The noise level at the output of the MSDW-MWF is defined as:

$$\begin{aligned} N \triangleq & \boldsymbol{w}^H \boldsymbol{\Phi}_{vv} \boldsymbol{w} \\ = & \|\boldsymbol{\Phi}_{vv}^{1/2} \boldsymbol{w}\|^2 \end{aligned} \tag{2.40}$$

where $\boldsymbol{\Phi}_{vv} = \left(\boldsymbol{\Phi}_{vv}^{1/2}\right)^H \boldsymbol{\Phi}_{vv}^{1/2}$ is the Cholesky decomposition of the noise correlation matrix. Substituting (2.32) in (2.40) yields:

$$N = \|\boldsymbol{\Phi}_{vv}^{1/2} \boldsymbol{\Psi} \left(\boldsymbol{I} + \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_{ss}^{-1} \boldsymbol{\Theta}\right)^{-1} \boldsymbol{g}\|^2. \tag{2.41}$$

In case that a low distortion of the SOIs is required, the following approximation can be obtained, by substituting (2.34) in (2.40):

$$N \approx \|\boldsymbol{\Phi}_{vv}^{1/2} \boldsymbol{\Psi} \left(\boldsymbol{I} - \boldsymbol{\Lambda}^{-1} \boldsymbol{\Phi}_{ss}^{-1} \boldsymbol{\Theta}\right) \boldsymbol{g}\|^2. \tag{2.42}$$

2.2.2.3 The LCMV-BF special case

In this section, we show that the LCMV-BF is a special case of the MSDW-MWF. Consider the MSDW-MWF formula in (2.21), and the following choice of $\boldsymbol{\Lambda} \triangleq \mu^{-1} \boldsymbol{\Phi}_{ss}^{-1}$. Substituting the latter choice of $\boldsymbol{\Lambda}$ in (2.21) yields:

$$\begin{aligned} \boldsymbol{w} &= \left(\mu^{-1} \boldsymbol{H} \boldsymbol{H}^H + \boldsymbol{\Phi}_{vv}\right)^{-1} \mu^{-1} \boldsymbol{H} \boldsymbol{g} \\ &= \left(\boldsymbol{H} \boldsymbol{H}^H + \mu \boldsymbol{\Phi}_{vv}\right)^{-1} \boldsymbol{H} \boldsymbol{g}. \end{aligned} \tag{2.43}$$

By applying the Woodbury identity to (2.43) and after some manipulation we obtain:

$$\begin{aligned} \boldsymbol{w} = & \Bigg(\mu^{-1} \boldsymbol{\Phi}_{vv}^{-1} \boldsymbol{H} - \mu^{-1} \boldsymbol{\Phi}_{vv}^{-1} \boldsymbol{H} \\ & \times \left(\boldsymbol{I} + \mu \left(\boldsymbol{H}^H \boldsymbol{\Phi}_{vv}^{-1} \boldsymbol{H}\right)^{-1}\right)^{-1} \Bigg) \boldsymbol{g}. \end{aligned} \tag{2.44}$$

Assuming that μ is "small" such that $\|\mu\left(\boldsymbol{H}^H\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{H}\right)^{-1}\| \ll 1$, we can replace $\left(\boldsymbol{I}+\mu\left(\boldsymbol{H}^H\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{H}\right)^{-1}\right)^{-1}$ by its first order Taylor series expansion $\boldsymbol{I}-\mu\left(\boldsymbol{H}^H\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{H}\right)^{-1}$. Finally, by substituting the latter approximation in (2.44), the LCMV-BF which satisfies the constraint set $\boldsymbol{H}^H\boldsymbol{w}=\boldsymbol{g}$ is obtained:

$$\boldsymbol{w}\approx\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{H}\left(\boldsymbol{H}^H\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{H}\right)^{-1}\boldsymbol{g}. \tag{2.45}$$

2.2.2.4 A modified MSDW-MWF

In practice, it is a cumbersome task to estimate the ATFs, $\boldsymbol{H}$, and the covariance matrix of the SOIs, $\boldsymbol{\Phi}_{ss}$. In this section, we obtain a modified MSDW-MWF which makes use of the RTFs and the covariance matrix of SOIs as received by some reference microphone. Without loss of generality let us define the RTFs of the SOIs with respect to the first microphone. Define the RTF of the pth source by $\tilde{\boldsymbol{h}}_p=\frac{\boldsymbol{h}_p}{h_{p,1}}$, and the RTF matrix by:

$$\tilde{\boldsymbol{H}}\triangleq\left[\begin{array}{ccc}\tilde{\boldsymbol{h}}_1 & \cdots & \tilde{\boldsymbol{h}}_P\end{array}\right]. \tag{2.46}$$

Next, we redefine the SOIs as their respective components in the first microphone, i.e., the pth modified SOI is given by $\tilde{s}_p(\ell)=h_{p,1}s_p(\ell)$, for $p=1,\ldots,P$. The corresponding modified SOIs covariance matrix equals

$$\tilde{\boldsymbol{\Phi}}_{ss}\triangleq\operatorname{diag}\left\{|h_{1,1}|^2\phi_{ss,1},\ldots,|h_{P,1}|^2\phi_{ss,P}\right\}. \tag{2.47}$$

Finally, substituting (2.46) and (2.47) in (2.21) yields the modified MSDW-MWF:

$$\tilde{\boldsymbol{w}}=\left(\tilde{\boldsymbol{H}}\boldsymbol{\Lambda}\tilde{\boldsymbol{\Phi}}_{ss}\tilde{\boldsymbol{H}}^H+\boldsymbol{\Phi}_{vv}\right)^{-1}\tilde{\boldsymbol{H}}\boldsymbol{\Lambda}\tilde{\boldsymbol{\Phi}}_{ss}\boldsymbol{g}. \tag{2.48}$$

For estimating the RTFs we use a similar subspace based procedure as in [51], and for the estimation the SOIs covariance matrix, we use a spectral substraction technique as in [30].

2.2.3 Experimental study

Here, we examine the MSDW-MWF in the case of a wideband stationary noise and two speech SOIs, a desired speaker and an interfering speaker. The dimensions of the simulated room are 4m $\times$ 3m $\times$ 3m, the reverberation time is set to 0.3s, an array of 8 microphones is located next to one of the walls and the sampling rate is set to 8kHz. The input SIR and SNR

levels are set to 0dB and 13dB, respectively, and the measured performance criteria are the output SIR, the NR, defined as the ratio $\frac{N}{\Phi_{vv}(1,1)}$ and the distortion levels of the SOIs d_1 and d_2. Note that the second SOI is an interfering source, which we would like to suppress. The distortion criterion (2.38) is suitable also for interfering sources. We would like to emphasize its meaning for this case. Usually, the desired response of an interfering SOI is zero, therefore the distortion is actually the power ratio of the interference at the output and the input. Explicitly, lower distortion means higher suppression. The performance is measured for various values of desired distortion levels, $\dot{d}_1$ and $\dot{d}_2$, in the range $[-25\text{dB}, -20\text{dB}, \ldots, 0\text{dB}]$. For each pair of $\dot{d}_1$, $\dot{d}_2$, the performance figures of merit are averaged over 20 Monte-Carlo experiments, in which the locations of sources are randomly selected. We use a window size of 4096 samples with 75% overlap. The distortions of the desired and interfering sources are depicted in Figs. 2.1,2.2, respectively. Clearly, from these figures, the MSDW-MWF allows for controlling individual distortion levels of the various SOIs. Note that the distortion level of the desired source, d_1 is lower bounded by -15dB, due to estimation errors of the RTF. As the approximated distortion levels (2.37) are valid for low distortion, the measured distortion on the interfering source in Fig. 2.2 differs from the desired one for higher levels of distortion $\dot{d}_1$. The NR versus the SOIs desired distortions is depicted in Fig. 2.3. This figure exemplifies that the NR can be controlled by sacrificing the distortion of just a sub-group of the SOIs. The average output SIR versus the desired SOIs distortion levels is depicted in Fig. 2.4. Note, that since the desired response of the desired source is 1, the variation in its output power is small for desired distortion levels $\dot{d}_1 \ll 0$dB. Therefore, as evidently seen in this figure, the output SIR is mainly determined by the desired distortion level of the interfering source.

2.2.4 Conclusion

We have considered the multiple SOIs in a noisy and reverberant environment scenario and extended the SDW-MWF for this case. The proposed method, denoted MSDW-MWF, allows for a better control of the tradeoff between NR and distortion levels of SOIs. We derive the SDW-MWF and the LCMV-BF as two special cases of the proposed method. We analyze the distortion levels of the various SOIs as well as the NR, and derive a more compact and simple approximation for the latter figures of merit, in the case of designing a low distortion MSDW-MWF. Finally, we exemplify the extended control over the NR versus distortion tradeoff in an experimental study.

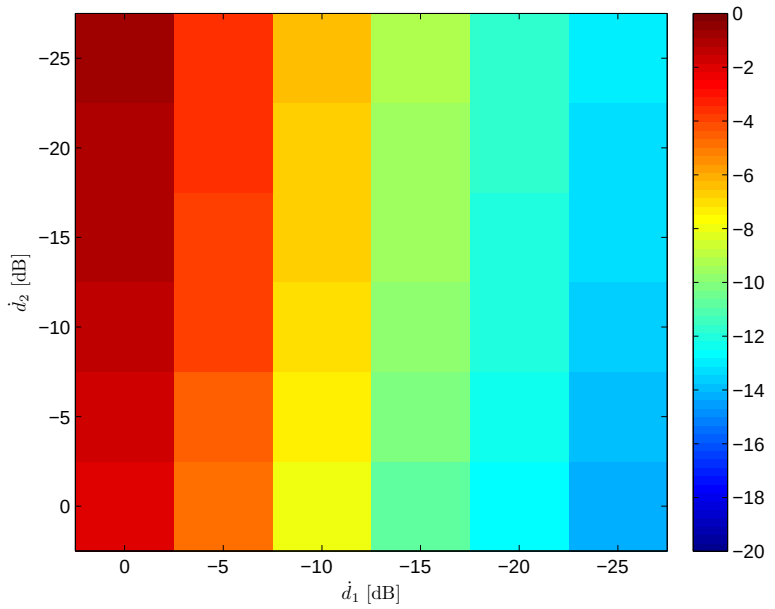


Figure 2.1: The distortion of the desired source, d_1, versus the $\dot{d}_1$, $\dot{d}_2$.

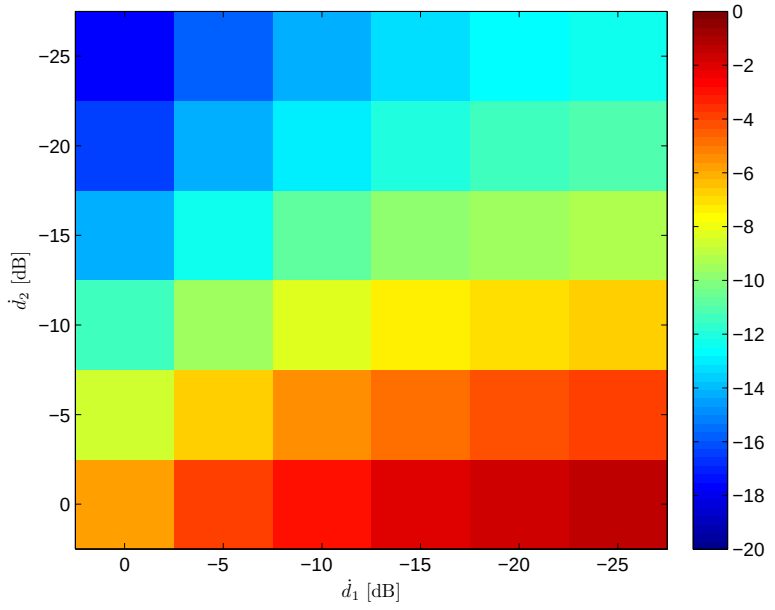


Figure 2.2: The distortion of the interfering source, d_2, versus $\dot{d}_1$, $\dot{d}_2$.

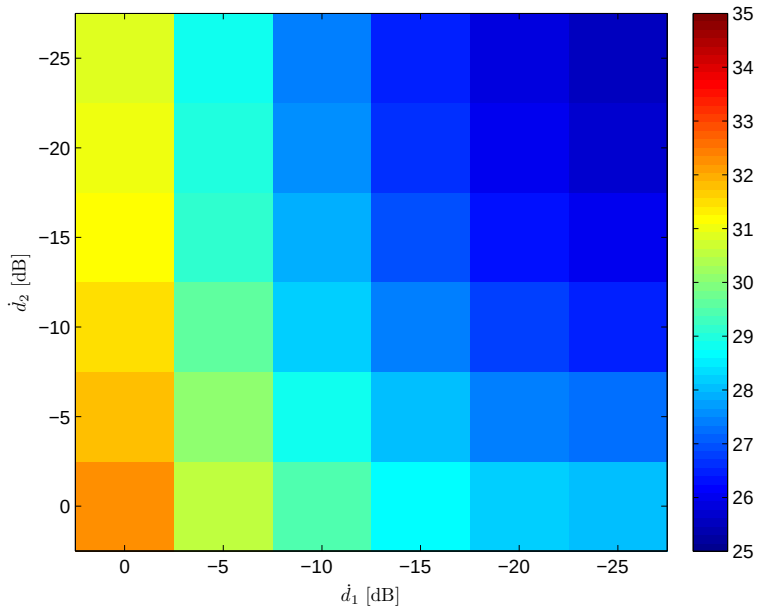


Figure 2.3: The NR versus the desired distortion levels $\dot{d}_1$, $\dot{d}_2$.

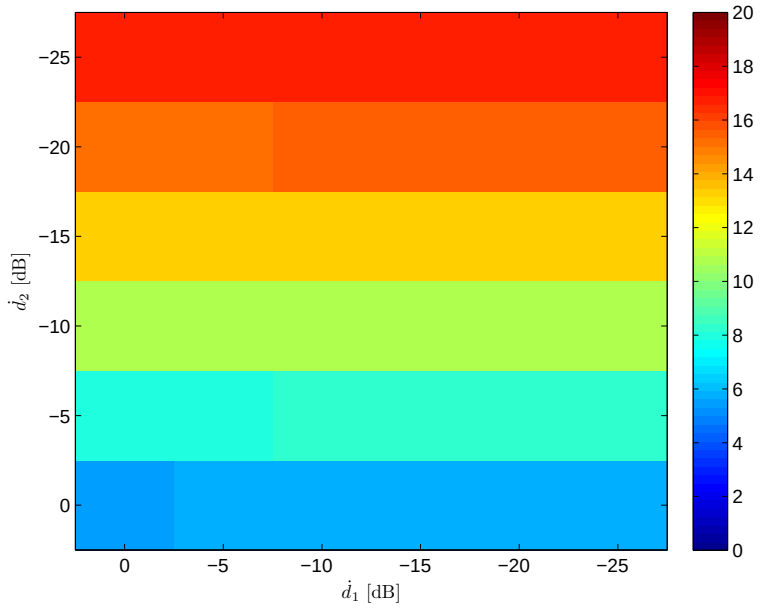


Figure 2.4: The output SIR versus the desired distortion levels $\dot{d}_1$, $\dot{d}_2$.

2.3 Subspace tracking of multiple sources and its application to speakers extraction

In Sec. 2.1 we introduced the eigen-space LCMV. BF for extracting the desired signals from multi-microphone measurements in a static scenario. The BF satisfies two sets of linear constraints. One set is dedicated to maintaining the desired signals, while the other set is chosen to mitigate both the stationary and non-stationary interferences.

The proposed algorithm however is inappropriate for dynamic scenarios. In practice, the ATFs of the speakers are constantly changing, due to slight movements of the sources or the reflectors. Affes et al. [46] constructed a GSC beamformer for the multi-source dynamic scenario. The proposed algorithm is based on the PASTd algorithm [43] for tracking the signals' subspace and on the MUSIC algorithm for estimating the steering vectors of the sources. The far-field regime and reverberation free environment allow tracking of the steering vectors during multi-speaker scenarios. However, its performance in reverberant scenarios is limited. Warsitz and Haeb-Umbach [103] used an alternative tracking procedure, based on the gradient ascent method, applied directly to the beamformer filters.

In this section, a novel algorithm for extracting desired speech signals uttered by moving speakers contaminated by competing speakers and stationary noise in a reverberant environment is presented. The proposed BF uses eigenvectors spanning the desired and interference signals subspaces. It relaxes the common requirement on the activity patterns of the various sources. A novel mechanism for tracking the desired and interferences subspaces is proposed, based on the PASTd procedure and on a union of subspaces procedure. This contribution extends previously proposed methods to deal with multiple speakers in dynamic scenarios.

The structure of this section is as follows. In Sec. 2.3.1 we formulate the speakers extraction problem. In Sec. 2.3.2 we extend the estimation algorithm described in Sec. 2.1 and in [51], and use an arbitrary subspace spanning the desired ATFs (instead of the RTFs). This relaxes the common requirement for non overlapping activity patterns of the desired sources. In Sec. 2.3.3 we introduce a novel mechanism for tracking the desired and interferences subspaces in a reverberant environment. The proposed speakers extraction algorithm is tested in both simulated and real environments in Sec. 2.3.4.

2.3.1 Problem Formulation

Consider the problem of extracting N_d desired speech signals $s_1^d(n),\ldots,s_{N_d}^d(n)$ uttered by moving speakers contaminated by N_i competing moving speakers $s_1^i(n),\ldots,s_{N_i}^i(n)$ as well as stationary interferences in a reverberant environment. Each of the involved signals undergo filtering before being picked up by M microphones arranged in an arbitrary array. The reverberation effect can be modeled by a finite impulse response (FIR) time-varying filtering. The received signals can be formulated in a vector notation, in the STFT domain as $\boldsymbol{z}(\ell,k) = \boldsymbol{H}^d(\ell,k)\boldsymbol{s}^d(\ell,k)+\boldsymbol{H}^i(\ell,k)\boldsymbol{s}^i(\ell,k)+\boldsymbol{v}(\ell,k)$ where $\boldsymbol{s}^d(\ell,k) = \begin{bmatrix} s_1^d(\ell,k) & \cdots & s_{N_d}^d(\ell,k) \end{bmatrix}^T$ and $\boldsymbol{s}^i(\ell,k) = \begin{bmatrix} s_1^i(\ell,k) & \cdots & s_{N_i}^i(\ell,k) \end{bmatrix}^T$ are vectors comprising the desired and interfering speech signals, respectively. k denotes the frequency index and ℓ the frame index. $\boldsymbol{H}^d(\ell,k) = \begin{bmatrix} \boldsymbol{h}_1^d(\ell,k) & \cdots & \boldsymbol{h}_{N_d}^d(\ell,k) \end{bmatrix}$ and $\boldsymbol{H}^i(\ell,k) = \begin{bmatrix} \boldsymbol{h}_1^i(\ell,k) & \cdots & \boldsymbol{h}_{N_i}^i(\ell,k) \end{bmatrix}$ are $M\times N_d$ and $M\times N_i$ matrices that involve time-varying ATFs relating the desired and interfering sources and the microphone array. $\boldsymbol{v}(\ell,k)$ denotes stationary noise components of the received signals, consisting of directional as well as spatially white signals.

Assuming the sources and the noise signals are uncorrelated, the correlation matrix of the received signals can be written as:

$$\begin{aligned}\boldsymbol{\Phi}_{zz}(\ell,k) = \boldsymbol{H}^d(\ell,k)\boldsymbol{\Lambda}^d(\ell,k)\big(\boldsymbol{H}^d(\ell,k)\big)^H \\ +\boldsymbol{H}^i(\ell,k)\boldsymbol{\Lambda}^i(\ell,k)\big(\boldsymbol{H}^i(\ell,k)\big)^H+\boldsymbol{\Phi}_{vv}(\ell,k)\end{aligned} \tag{2.49}$$

where $\boldsymbol{\Lambda}^d(\ell,k) \triangleq \operatorname{diag}\left(\begin{bmatrix} (\sigma_1^d(\ell,k))^2 & \ldots & (\sigma_{N_d}^d(\ell,k))^2 \end{bmatrix}\right)$ and $\boldsymbol{\Lambda}^i(\ell,k) \triangleq \operatorname{diag}\left(\begin{bmatrix} (\sigma_1^i(\ell,k))^2 & \ldots & (\sigma_{N_i}^i(\ell,k))^2 \end{bmatrix}\right)$ are diagonal matrices with the spectral variances of the desired and interfering sources on their main diagonal respectively. $\boldsymbol{\Phi}_{vv}(\ell,k)$ is the stationary noise correlation matrix. $(\bullet)^H$ is the conjugate-transpose operation, and $\operatorname{diag}(\bullet)$ is a square matrix with the vector in brackets on its main diagonal. In the following section we derive an algorithm for extracting the desired sources while mitigating the interferences in dynamic environments.

2.3.2 Speakers Extraction in a Dynamic Environment

Markovich-Golan et al. [51] propose a novel eigenspace based LCMV beamformer, designed for extracting static desired sources. Rather than using the sources' ATFs for constructing the constraints set, they use an arbitrary basis for the interferences subspace and the RTFs

of the desired sources. They also derive an algorithm for estimating the subspace, that spans the non-stationary interference signals, having an arbitrary activity pattern.

Following [51], define a modified constraints set $\dot{\boldsymbol{C}}^H(\ell,k)\boldsymbol{w}(\ell,k)=\boldsymbol{g}(\ell,k)$ where $\dot{\boldsymbol{C}}(\ell,k)=\begin{bmatrix}\tilde{\boldsymbol{H}}^d(\ell,k) & \boldsymbol{Q}^i(\ell,k)\end{bmatrix}$ is the constraints matrix and $\boldsymbol{g}(\ell,k)\triangleq\begin{bmatrix}\underbrace{1\ \dots\ 1}_{N_d} & \underbrace{0\ \dots\ 0}_{N_i}\end{bmatrix}^T$ is the desired response vector. $\boldsymbol{Q}^i(\ell,k)$ denotes an orthonormal basis which spans the interferences subspace, i.e. $\boldsymbol{H}^i(\ell,k)=\boldsymbol{Q}^i(\ell,k)\boldsymbol{\Theta}^i(\ell,k)$ where $\boldsymbol{\Theta}^i(\ell,k)$ is the projection coefficients matrix. $\tilde{\boldsymbol{H}}^d(\ell,k)=\begin{bmatrix}\tilde{\boldsymbol{h}}^d_1(\ell,k) & \cdots & \tilde{\boldsymbol{h}}^d_{N_d}(\ell,k)\end{bmatrix}$ denotes a matrix of the desired sources' RTFs with respect to reference microphone #1. The RTF of the ith desired source is defined as $\tilde{\boldsymbol{h}}^d_i(\ell,k)=\frac{1}{h^d_{i1}(\ell,k)}\boldsymbol{h}^d_i(\ell,k)$. The closed form beamformer solving this problem is given by:

$$\begin{aligned}\boldsymbol{w}(\ell,k)=&\boldsymbol{\Phi}^{-1}_{zz}(\ell,k)\dot{\boldsymbol{C}}(\ell,k)\\&\times\left(\dot{\boldsymbol{C}}^H(\ell,k)\boldsymbol{\Phi}^{-1}_{zz}(\ell,k)\dot{\boldsymbol{C}}(\ell,k)\right)^{-1}\boldsymbol{g}(\ell,k).\end{aligned}\tag{2.50}$$

They further propose the use of the QRD procedure to perform the union of basis vectors obtained from several time segments. Estimating the constraint matrix utilizes segments of simultaneously active interference sources, but discards segments of desired signals' double-talk. In the sequel, we further relax the latter requirement, allowing simultaneously active desired sources in the estimation procedure.

Denote by $\boldsymbol{Q}^d(\ell,k)$ an orthonormal basis spanning the desired subspace $\boldsymbol{H}^d(\ell,k)=\boldsymbol{Q}^d(\ell,k)\boldsymbol{\Theta}^d(\ell,k)$ where $\boldsymbol{\Theta}^d(\ell,k)$ is the projection coefficients matrix. We propose to use the following modified constraints set

$$\tilde{\boldsymbol{C}}(\ell,k)=\begin{bmatrix}\boldsymbol{Q}^d(\ell,k) & \boldsymbol{Q}^i(\ell,k)\end{bmatrix}\tag{2.51a}$$

$$\tilde{\boldsymbol{g}}(\ell,k)=\begin{bmatrix}\underbrace{\left(Q^d_{11}(\ell,k)\right)^*\ \dots\ \left(Q^d_{N_d1}(\ell,k)\right)^*}_{N_d} & \underbrace{0\ \dots\ 0}_{N_i}\end{bmatrix}^T\tag{2.51b}$$

where we substitute the desired sources' RTFs in the constraints matrix $\dot{\boldsymbol{C}}(\ell,k)$ by the basis $\boldsymbol{Q}^d(\ell,k)$, and $\begin{bmatrix}1 & \cdots & 1\end{bmatrix}_{1\times N_d}$ in the desired response vector $\boldsymbol{g}(\ell,k)$ by the first row of $\boldsymbol{Q}^d(\ell,k)$.

Let $\tilde{\boldsymbol{w}}(\ell,k)$ be the solution of the LCMV with the modified constraints set. The output

of the modified beamformer is given by:

$$\tilde{y}(\ell,k) = \tilde{\boldsymbol{w}}^H(\ell,k)\boldsymbol{z}(\ell,k) = \sum_{j=1}^{N_d} h_{j1}^d(\ell,k)s_j^d(\ell,k) + \tilde{\boldsymbol{g}}^H(\ell,k)\big(\tilde{\boldsymbol{C}}^H(\ell,k)\boldsymbol{\Phi}_{vv}^{-1}(\ell,k)\tilde{\boldsymbol{C}}(\ell,k)\big)^{-1}\tilde{\boldsymbol{C}}^H(\ell,k)\boldsymbol{v}(\ell,k). \tag{2.52}$$

Hence, the desired sources as received by the reference microphone are extracted, the non-stationary interferences are mitigated, and the power of the remaining stationary noise is minimized. Although the union based subspace estimation method obtains good performance with static sources, it is rendered useless when they are allowed to move, since the rank of the estimated subspace may excessively grow. Without prior knowledge of the rank, source movement, manifested as ATF change, results in a *birth* of a new basis vector. We circumvent this phenomenon by incorporating a *death* mechanism for the obsolete basis vectors in the estimation procedure. A novel subspace tracking algorithm utilizing birth and death mechanism is introduced in the following section.

2.3.3 Proposed Subspace Tracking Algorithm

The proposed tracking algorithm is based on the classic PASTd procedure introduced by Yang [43]. The PASTd procedure is a recursive algorithm incorporating a forgetting factor β. The latter results in an inherent memory of $N_\beta \approx \frac{1}{1-\beta}$ frames, contributing to the subspace estimation. The main limitation in applying the PASTd procedure to the problem at hand stems from conflicting memory requirements. On the one hand, we would like to apply PASTd with short memory in order to have fast adaptation time, and to quickly react to birth or death of basis vectors. On the other hand, using short memory, only recently active speakers will be included in the estimated subspace. All other speakers effectively die out. As a consequence, during the adaptation time, desired speakers that resume activity might suffer distortion, and competing speakers that resume activity may not be canceled out.

We propose to settle these contradicting requirements by using a short memory PASTd, allowing for fast adaption of basis vectors. Yet, basis vectors meeting certain conditions are declared *stable* and remain part of the estimated subspace for a predefined *expiry-time*. The stability conditions are explained in Sec. 2.3.3.2.

The proposed subspace tracking algorithm consists of three stages. First, a generalized PASTd procedure tracks the current subspace as explained in Sec. 2.3.3.1. Second, the expiry

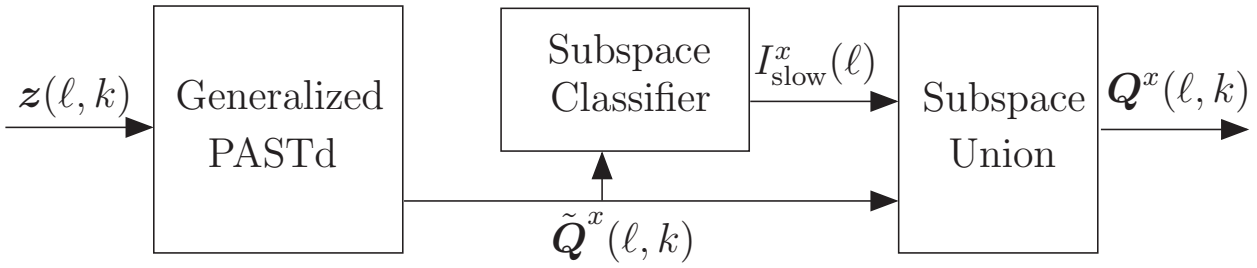


Figure 2.5: Block diagram of the proposed tracking algorithm.

time is attributed to stable basis vectors. Third, the current basis vectors and the valid stable basis vectors are combined by using the union operation as explained in Sec. 2.3.3.3. A block diagram of the proposed tracking scheme is depicted in Fig. 2.5.

2.3.3.1 PASTd – Subspace Tracking

As we are dealing with two distinct groups of signals (desired and interfering) we apply the tracking algorithm to each group independently. Note that the proposed subspace tracking algorithm can only operate on time-segments in which desired and interfering speakers are mutually inactive. It is assumed that these segments exist and they are used for tracking the respective signal subspaces. Let x denote the active group, where $x \in \{d, i\}$. Define the activity indicator of the xth group

$$I^x(\ell) = \begin{cases} 1 & \text{only sources of the } x\text{th group are active} \\ 0 & \text{otherwise} \end{cases} . \tag{2.53}$$

We assume that this activity indicator is available to the algorithm. Note, that a group x is declared active if at least one of its signals is active. The activity indicator $I^x(\ell)$ is regulating the subspace tracking algorithm.

PASTd estimation method is only suitable for tracking the signal subspace in a spatially white noise environment. Therefore, a whitening procedure should precede the activation of the tracking algorithm. Denote the whitened microphone signals as $\boldsymbol{z}_w(\ell,k) = \boldsymbol{\Phi}^{-1}_{vv,L}(\ell,k)\boldsymbol{z}(\ell,k)$, where $\boldsymbol{\Phi}_{vv,L}(\ell,k)$ is the lower triangular matrix obtained by the Cholesky decomposition of the stationary noise covariance matrix, $\boldsymbol{\Phi}_{vv}(\ell,k) = \boldsymbol{\Phi}_{vv,L}(\ell,k)\boldsymbol{\Phi}^H_{vv,L}(\ell,k)$. The noise covariance matrix $\boldsymbol{\Phi}_{vv}(\ell,k)$ can be estimated by any conventional noise estimation procedure. The resulting covariance matrix of the whitened microphone signals is therefore

given by:

$$\boldsymbol{\Phi}_{z_w z_w}(\ell,k) = \boldsymbol{\Phi}_{vv,L}^{-1}(\ell,k)\boldsymbol{\Phi}_{zz}(\ell,k)\left(\boldsymbol{\Phi}_{vv,L}^{-1}(\ell,k)\right)^H. \tag{2.54}$$

The PASTd procedure tracks $N_u \leq M$ major eigenvectors of the two groups of the whitened sources $\{\boldsymbol{u}_r^x(\ell,k)\}_{r=1}^{N_u}$ and their corresponding eigenvalues $\{d_r^x(\ell,k)\}_{r=1}^{N_u}$. It is proven by Yang [43] that the estimated subspace converges to an orthonormal basis of the signals subspace.

A basis that spans the signal subspace of the original measurements $\boldsymbol{z}(\ell,k)$ is given by:

$$\tilde{\boldsymbol{u}}_r^x(\ell,k) = \sqrt{d_r^x(\ell,k)}\boldsymbol{\Phi}_{vv,L}(\ell,k)\boldsymbol{u}_r^x(\ell,k) \tag{2.55}$$

where we scaled the basis vectors by their corresponding eigenvalues.

Note that this representation is no longer orthogonal. To obtain an orthogonal representation the following steps are applied. Define, $\tilde{\boldsymbol{U}}^x(\ell,k) \triangleq \left[\begin{array}{ccc} \tilde{u}_1^x(\ell,k) & \cdots & \tilde{u}_{N_u}^x(\ell,k) \end{array}\right]$. Next, a QRD is applied to $\tilde{\boldsymbol{U}}^x(\ell,k)$. Finally, the required orthogonal basis $\tilde{\boldsymbol{Q}}^x(\ell,k)$ is obtained by selecting the dominant vectors spanning $\tilde{\boldsymbol{U}}^x(\ell,k)$ scaled by their corresponding energy.

2.3.3.2 Classification of Subspace Stability

The basis $\tilde{\boldsymbol{Q}}^x(\ell,k)$ defined in the previous section spans the subspace of the currently active sources in group x. Recall that this basis is always valid for at least N_β frames, due to the inherent memory of the PASTd technique. In a static scenario these basis vectors should remain unaltered. Based on this property, we propose a classification criterion for subspace stability. We define an indicator function

$$I_{\text{stable}}^x(\ell) \triangleq \begin{cases} 1 & \{\tilde{\boldsymbol{Q}}^x(\ell,k)\}_{k=0}^{N_{\text{DFT}}-1} \text{ is stable} \\ 0 & \text{otherwise} \end{cases}.$$

Each subspace that is valid for more than N_{stable} frames will be declared stable. Define the projection matrix to $\tilde{\boldsymbol{Q}}^x(\ell,k)$, the signal subspace, by

$$\boldsymbol{P}_{\tilde{Q}^x}(\ell,k) \triangleq \tilde{\boldsymbol{Q}}^x(\ell,k)\left(\left(\tilde{\boldsymbol{Q}}^x(\ell,k)\right)^H\tilde{\boldsymbol{Q}}^x(\ell,k)\right)^{-1}\left(\tilde{\boldsymbol{Q}}^x(\ell,k)\right)^H. \tag{2.56}$$

The energy of the projection of the received signals in frame ℓ' to the current basis $\{\tilde{\boldsymbol{Q}}^x(\ell,k)\}_{k=0}^{N_{\mathrm{DFT}}-1}$ is given by:

$$E_{\tilde{Q}^x}(\ell',\ell) \triangleq \sum_{k=0}^{N_{\mathrm{DFT}}-1} \alpha^x(\ell,k)\|\boldsymbol{P}_{\tilde{Q}^x}(\ell,k)\boldsymbol{z}(\ell',k)\|^2 \tag{2.57}$$

where $\alpha(\ell,k) \triangleq 1 - \frac{N_{\tilde{Q}^x}(\ell,k)}{M}$ is a compensation factor for high signal subspace rank. Hence, the aggregated projection energy over N_{stable} frames is given by:

$$E_{\tilde{Q}^x}(\ell) = \sum_{j=0}^{N_{\mathrm{stable}}-1} E_{\tilde{Q}^x}(\ell - N_\beta - j, \ell).$$

Finally, we set $I^x_{\mathrm{stable}}(\ell) = 1$ if $\frac{E_{\tilde{Q}^x}(\ell)}{E^x(\ell)}$ is higher than a predefined threshold, where $E^x(\ell)$ is the aggregated energy of the received signal over N_{stable} frames. Subspaces that are declared stable are attributed with an expiry-time. The expiry-time provides a mechanism for forgetting unused basis vectors.

2.3.3.3 Subspaces Union

To guarantee that basis vectors common to the current subspace and the stable subspaces are not counted more than once they should be collected by the union operator (see an analogue discussion in [51]). The union operator can be implemented in many ways. Here we chose to use the QRD. The required orthonormal basis $\boldsymbol{Q}^x(\ell,k)$ for group x is obtained by selecting the dominant vectors spanning the collection of valid subspaces. Note that the rank of the signal subspace is estimated from the received data and therefore the knowledge of N_d and N_i is not required.

2.3.4 Experimental Study

The proposed algorithm is tested with simulated signals as well as with real signals recorded in our acoustics lab. We examine a scenario in which two desired speakers and two interfering speakers are moving around in a reverberant noisy environment. The dimensions of the simulated room are $3\mathrm{m} \times 4\mathrm{m} \times 2.7\mathrm{m}$. The reverberation time is set to 0.3s in both environments. The acoustics lab and the simulated room are depicted in Figs. 2.6, 2.7, respectively. The microphone array comprises 9 microphones and is arranged in a non-uniform linear array with total length of 0.64m. The SIR (with respect to the non-stationary interferences) and

SNR (with respect to the stationary interference) are 0dB and 30dB, respectively. The sonogram and the waveform of the signal received by a reference microphone, in the acoustics lab scenario, are depicted in Fig. 2.8(a). The respective output of the proposed algorithm is depicted in Fig. 2.8(b). Comparing both signals, it is clearly seen that the interference signals are significantly attenuated, especially in high frequency bands. The SIR improvement in the acoustics lab, using the proposed algorithm, is 7.5dB, while in the simulated environment is 9.7dB.

Figure 2.6: The acoustics lab at Bar-Ilan University premises.

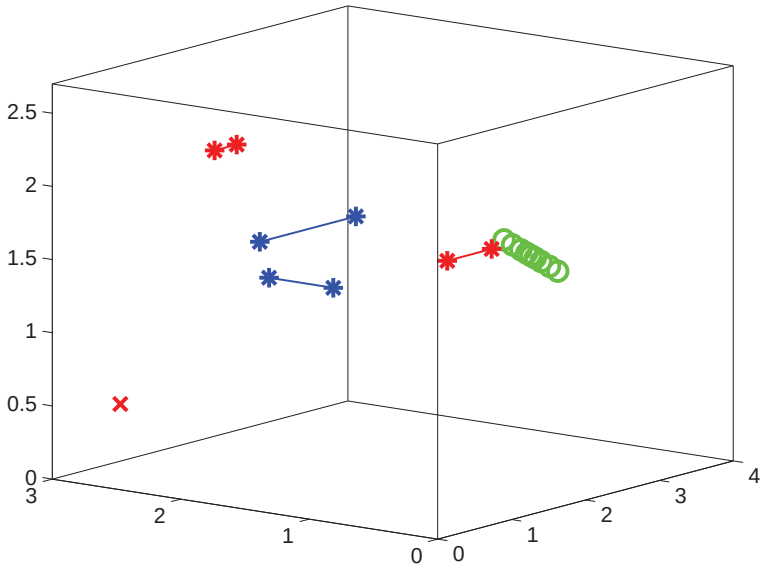
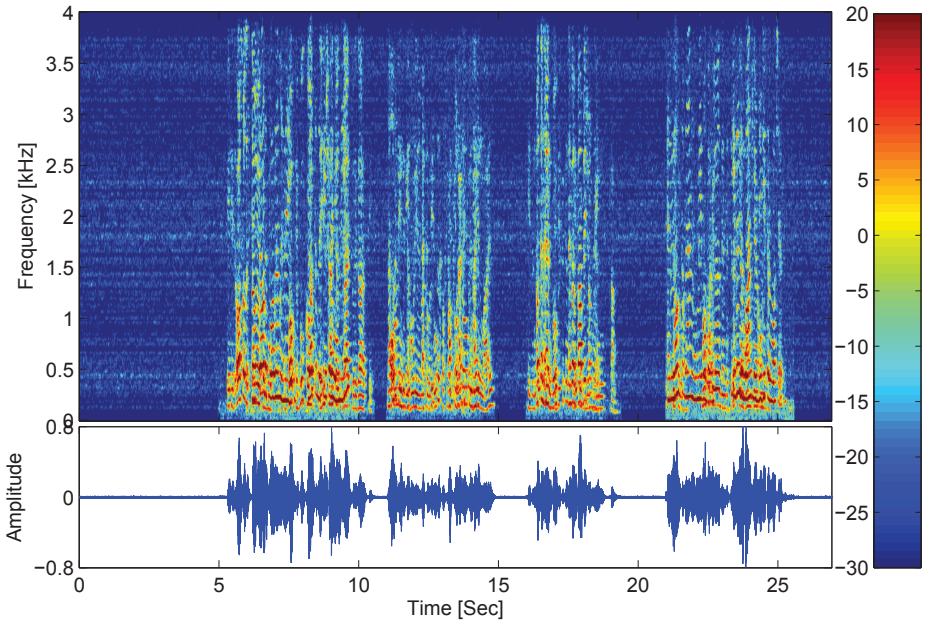
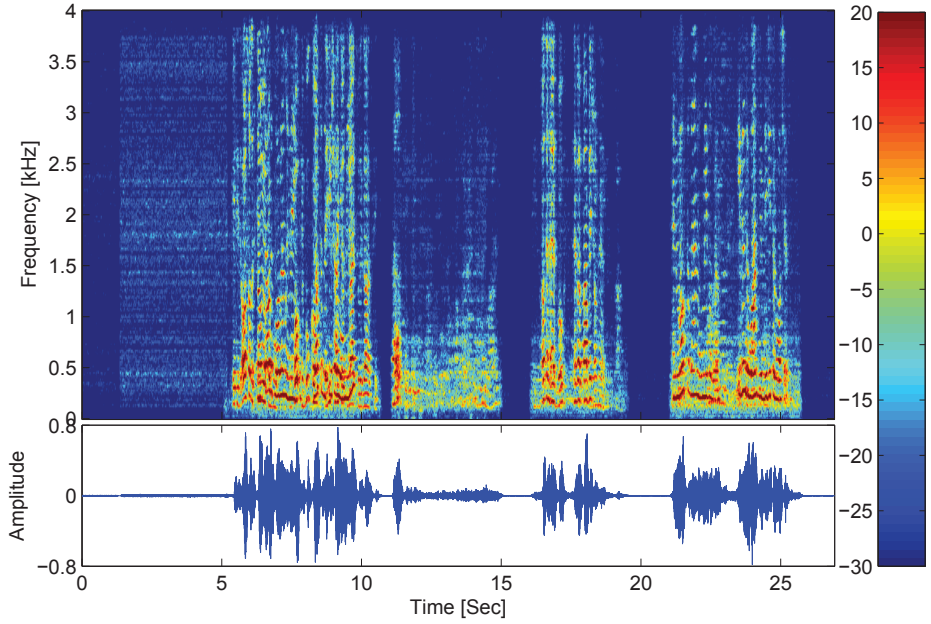


Figure 2.7: The simulated scenario: Green circles denote the microphones. Blue and red stars denote desired and interfering sources, respectively. A line connecting two stars denotes the route of the source's movement. A red $\times$ denotes a stationary interference.



(a) Signal received by a reference microphone.



(b) The output of the beamformer.

Figure 2.8: Received signal and the beamformer output in a real environment with moving sources.

2.3.5 Conclusions

A novel algorithm for tracking signal subspaces has been introduced. The algorithm tracks the current subspace using the PASTd algorithm and classifies certain subspaces as stable. An expiry-time is then attributed to the stable subspaces. The union operator implemented by the QRD is used for collecting valid basis vectors, independently for the desired and the interfering groups of signals. The resulting signals subspaces are used to construct a BF for extracting desired sources in a dynamic environment. The proposed tracking algorithm relaxes limiting requirements on sources activity (common to other algorithms), and allows for simultaneous source activity within the groups. The novel algorithm is shown to yield good results, both in real and simulated environments.

Chapter 3

Distributed beamforming

In this chapter, novel distributed BF algorithms based on well-known criteria are developed for WASNs. The distributed versions alleviate the communication bandwidth requirements of sharing all microphone data in the network. Iterative, and time-recursive procedures are derived, in which only partial data is transmitted. These procedures are proven to converge to the optimal centralized BF over time.

In Sec. 3.1, a distributed MVDR-BF is derived for binaural hearing aid systems. The latter is a simple WASN topology which comprises two nodes, the left and right hearing apparatuses. Considering a scenario of a single desired speaker in a noisy environment, we propose a distributed binaural MVDR-BF which requires only a single bidirectional communication channel with a bandwidth of a single microphone. The proposed algorithm is iterative, and converges to the centralized solution after several iterations.

In Sec. 3.2, the multiple speakers scenario is addressed. A distributed GSC-BF, which reduces the required communication bandwidth significantly, without sacrificing performance, is suggested. Denoting N, the number of nodes, and P, the number of speakers, the required communication bandwidth is equal to $N + P$ audio channels.

In Sec. 3.3, we address the more simple scenario of a single desired speaker in a noisy environment received by a WASN. We develop a GSC-BF for this special case which requires only N transmission channels, rather than $N + 1$ when the distributed BF of Sec. 3.2 is applied.

All distributed BFs in this chapter assume that the microphone signals in all nodes are sampled with an equal and accurate sampling rate. However, this is hardly ever the case as each node might use an independent clock source and sampling rate offsets are inevitable.

In Sec. 3.4 we propose a blind synchronization algorithm for estimating the sampling rate offsets in a WASN and propose methods for compensating this offset. The procedure assumes that a noise with a coherent component and slowly time-varying statistics exists. Noise-only time segments are used to track coherence changes between microphone of various nodes, and hence to estimate the sampling-rate offsets. This procedure can be incorporated in the previously presented distributed BFs.

3.1 A reduced bandwidth binaural MVDR beamformer

A novel reduced bandwidth iterative algorithm for a distributed MVDR beamformer with application to binaural hearing aids is presented. The proposed algorithm requires a single transmission channel between laterals. It is well-known that the MVDR beamformer is a special case of the SDW-MWF. The convergence of the iterative procedure to the binaural MVDR is proved for a desired source correlation matrix with rank-1. The proposed method is shown to outperform the monaural MVDR, where the output is generated by filtering only local sensors, without communicating with the other side.

Maintaining spatial cues is a desired property of any binaural BF. The binaural MVDR maintains the spatial cues of the desired source as can be deduces from its distortionless response. However, spatial cues of the interfering sources are not maintained.

This section is organized as follows. In Sec. 3.1.1, the binaural hearing aids problem is formulated. In Sec. 3.1.2, a closed-form solution for the binaural MVDR based on all sensors' data is derived. In Sec. 3.1.3, the novel reduced bandwidth iterative MVDR algorithm is proposed. The proof of convergence in the rank-1 scenario is presented in the Sec. 3.1.4. Finally, the algorithm is evaluated for narrow-band stationary signals, and for speech signals in reverberant environments in Sec. 3.1.5.

3.1.1 Problem Formulation

The problem is formulated in the STFT domain. Consider a desired speech signal $s(\ell, k)$ impinging on two microphone arrays in the left and right hearing aid apparatuses placed in a reverberant environment. The received signals are contaminated by a stationary noise $\boldsymbol{v}(\ell, k)$. From here on we omit the time and frequency indexes for brevity. The signals received by

the left and right arrays are given by $\boldsymbol{z}_l = \boldsymbol{h}_l s + \boldsymbol{v}_l$ and $\boldsymbol{z}_r = \boldsymbol{h}_r s + \boldsymbol{v}_r$, respectively, where $\boldsymbol{h}_l, \boldsymbol{h}_r$ are the ATFs relating the desired source and the left and right arrays. Define the vectors comprised of a concatenation of the left and right signals $\boldsymbol{z} = \begin{bmatrix} \boldsymbol{z}_l^T & \boldsymbol{z}_r^T \end{bmatrix}^T = \boldsymbol{h}s + \boldsymbol{v}$, where $\boldsymbol{v} = \begin{bmatrix} \boldsymbol{v}_l^T & \boldsymbol{v}_r^T \end{bmatrix}^T$ and $\boldsymbol{h} = \begin{bmatrix} \boldsymbol{h}_l^T & \boldsymbol{h}_r^T \end{bmatrix}^T$. Denote the covariance matrix of the received signals:

$$\boldsymbol{\Phi}_{zz} = \sigma^2 \boldsymbol{h}^H \boldsymbol{h} + \boldsymbol{\Phi}_{vv} = \begin{bmatrix} \boldsymbol{\Phi}_{ll} & \boldsymbol{\Phi}_{lr} \\ \boldsymbol{\Phi}_{rl} & \boldsymbol{\Phi}_{rr} \end{bmatrix} \tag{3.1}$$

where $\boldsymbol{\Phi}_{vv} = \mathrm{E}\left[\boldsymbol{v}\boldsymbol{v}^H\right]$ is the covariance matrix of the stationary noise. The goal of the standard binaural MVDR beamformer is to reduce the noise power at two reference microphones at the left and right apparatuses, by using all microphone data, and while keeping the desired speech components undistorted. In this section a distributed version of the binaural MVDR is addressed. The algorithm should limit communication bandwidth between laterals without sacrificing the performance.

3.1.2 Closed-form Binaural MVDR

The binaural MVDR beamformer consists of two beamformers designed for reproducing the desired signal components as received by reference microphones in each lateral, while minimizing the overall noise power. The output signals of the closed-form beamformer are given by $y_l^o = (\boldsymbol{w}^{ol})^H \boldsymbol{z}$ and $y_r^o = (\boldsymbol{w}^{or})^H \boldsymbol{z}$. The closed-form solution is given by:

$$\boldsymbol{w}^{ol} = \left((\tilde{\boldsymbol{h}}^l)^H \boldsymbol{\Phi}_{zz}^{-1} \tilde{\boldsymbol{h}}^l \right)^{-1} \boldsymbol{\Phi}_{zz}^{-1} \tilde{\boldsymbol{h}}^l \tag{3.2a}$$

$$\boldsymbol{w}^{or} = \left((\tilde{\boldsymbol{h}}^r)^H \boldsymbol{\Phi}_{zz}^{-1} \tilde{\boldsymbol{h}}^r \right)^{-1} \boldsymbol{\Phi}_{zz}^{-1} \tilde{\boldsymbol{h}}^r \tag{3.2b}$$

where the left and right RTFs are defined as:

$$\tilde{\boldsymbol{h}}^l = (h_{l,1})^{-1} \boldsymbol{h} = \begin{bmatrix} (\tilde{\boldsymbol{h}}_l^l)^T & (\tilde{\boldsymbol{h}}_r^l)^T \end{bmatrix}^T$$

$$\tilde{\boldsymbol{h}}^r = (h_{r,1})^{-1} \boldsymbol{h} = \begin{bmatrix} (\tilde{\boldsymbol{h}}_l^r)^T & (\tilde{\boldsymbol{h}}_r^r)^T \end{bmatrix}^T.$$

The first microphones are arbitrarily chosen as the reference microphones. Throughout this section the subscript notations $(\cdot)_l$ and $(\cdot)_r$ are used for denoting the vector components corresponding to the left and right laterals, respectively. The superscript notations $(\cdot)^l$ and

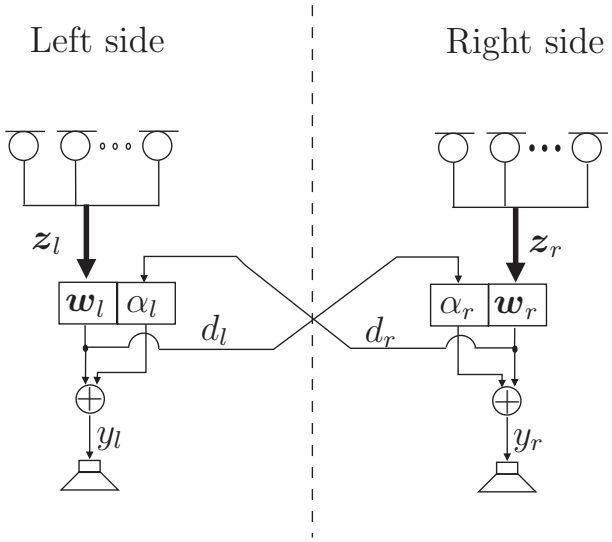


Figure 3.1: Diagram of the distributed binaural MVDR.

$(\cdot)^r$ are used to denote variables which are used for calculating the outputs of the left and right apparatuses, respectively. Note that the left and right MVDR beamformers are parallel in the rank-1 case.

3.1.3 Proposed method

In this section, we introduce a batch iterative algorithm which converges to the closed-form solution introduced in the previous section. We assume that the second moments of the observed data are available or can be estimated without errors. At each iteration, each side calculates the MVDR beamformer based on its local microphones and the channel received from the lateral side. Its contribution to the binaural beamformer is transmitted to the lateral side which in turn also updates its coefficients in a similar manner. Each iteration is therefore comprised of updating both left and right beamformers. A diagram of the algorithm is depicted in Fig. 3.1. Consider the ith iteration of the algorithm. Without loss of generality, we assume that the left side is the first to update its beamformer. The data available to the left side is its own microphones and the received channel from the right side in the previous iteration

$$d_r^{i-1} = (\boldsymbol{w}_r^{i-1})^H \boldsymbol{z}_r. \tag{3.3}$$

The MVDR equation at the ith iteration at the left side is given by

$$\begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \end{bmatrix} = \underset{\boldsymbol{w}_l^i, \alpha_l^i}{\operatorname{argmin}} \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \end{bmatrix}^H \mathrm{E}\left[\begin{bmatrix} \boldsymbol{z}_l \\ d_r^{i-1} \end{bmatrix} \begin{bmatrix} \boldsymbol{z}_l \\ d_r^{i-1} \end{bmatrix}^H \right] \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \end{bmatrix} \tag{3.4}$$
$$\text{s.t.} \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \end{bmatrix}^H \begin{bmatrix} \tilde{\boldsymbol{h}}_l^l \\ (\boldsymbol{w}_r^{i-1})^H \tilde{\boldsymbol{h}}_r^l \end{bmatrix} = 1.$$

The last minimization can be reformulated as a constrained minimization of the binaural MVDR, where the right coefficients are constant up to a scaling factor $\alpha_l^i \boldsymbol{w}_r^{i-1}$:

$$\begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \end{bmatrix} = \underset{\boldsymbol{w}_l^i, \alpha_l^i}{\operatorname{argmin}} \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \boldsymbol{w}_r^{i-1} \end{bmatrix}^H \boldsymbol{\Phi}_{zz} \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \boldsymbol{w}_r^{i-1} \end{bmatrix} \tag{3.5}$$
$$\text{s.t.} \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \boldsymbol{w}_r^{i-1} \end{bmatrix}^H \tilde{\boldsymbol{h}}^l = 1.$$

The minimization is performed by using Lagrange multipliers. Define the Lagrangian of the left side:

$$\begin{aligned} \mathcal{L}_l(\boldsymbol{w}_l^i, \alpha_l^i, \lambda_l^i) = & \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \boldsymbol{w}_r^{i-1} \end{bmatrix}^H \boldsymbol{\Phi}_{zz} \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \boldsymbol{w}_r^{i-1} \end{bmatrix} \\ & + \lambda_l^i \left(\begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \boldsymbol{w}_r^{i-1} \end{bmatrix}^H \tilde{\boldsymbol{h}}^l - 1 \right) \\ & + (\lambda_l^i)^* \left((\tilde{\boldsymbol{h}}^l)^H \begin{bmatrix} \boldsymbol{w}_l^i \\ \alpha_l^i \boldsymbol{w}_r^{i-1} \end{bmatrix} - 1 \right). \end{aligned} \tag{3.6}$$

Minimizing (3.6) is performed by solving the partial derivatives $\frac{\partial \mathcal{L}_l}{\partial (\boldsymbol{w}_l^i)^H} = \mathbf{0}$, $\frac{\partial \mathcal{L}_l}{\partial (\alpha_l^i)^*} = 0$, $\frac{\partial \mathcal{L}_l}{\partial (\lambda_l^i)^*} = 0$ and its solution is given by:

$$\lambda_l^i = \left(-\|\bar{\boldsymbol{h}}_l^{l,i}\|^2_{(\bar{\boldsymbol{\Phi}}_{ll}^i)^{-1}} - \frac{\|(\tilde{\boldsymbol{h}}_r^l)^H \boldsymbol{w}_r^{i-1}\|^2}{(\boldsymbol{w}_r^{i-1})^H \boldsymbol{\Phi}_{rr} \boldsymbol{w}_r^{i-1}} \right)^{-1} \tag{3.7a}$$

$$\boldsymbol{w}_l^i = \lambda_l^i (\bar{\boldsymbol{\Phi}}_{ll}^i)^{-1} \bar{\boldsymbol{h}}_l^{l,i} \tag{3.7b}$$

$$\alpha_l^i = - \frac{(\boldsymbol{w}_r^{i-1}) \boldsymbol{\Phi}_{rl} \boldsymbol{w}_l^i + \lambda_l^i (\boldsymbol{w}_r^{i-1})^H \tilde{\boldsymbol{h}}_r^l}{(\boldsymbol{w}_r^{i-1})^H \boldsymbol{\Phi}_{rr} \boldsymbol{w}_r^{i-1}} \tag{3.7c}$$

where

$$\bar{\boldsymbol{\Phi}}_{ll}^{i} = \boldsymbol{\Phi}_{ll} - \frac{\boldsymbol{\Phi}_{lr}\boldsymbol{w}_r^{i-1}(\boldsymbol{w}_r^{i-1})^H\boldsymbol{\Phi}_{rl}}{(\boldsymbol{w}_r^{i-1})^H\boldsymbol{\Phi}_{rr}\boldsymbol{w}_r^{i-1}} \tag{3.8a}$$

$$\bar{\boldsymbol{h}}_l^{l,i} = \frac{(\boldsymbol{w}_r^{i-1})^H\tilde{\boldsymbol{h}}_r^{\,l}\boldsymbol{\Phi}_{lr}\boldsymbol{w}_r^{i-1}}{(\boldsymbol{w}_r^{i-1})^H\boldsymbol{\Phi}_{rr}\boldsymbol{w}_r^{i-1}} - \tilde{\boldsymbol{h}}_l^{\,l} \tag{3.8b}$$

and the matrix norm with respect to the matrix $\boldsymbol{A}$ is denoted by $\|\cdot\|_{\boldsymbol{A}}$. Note that the second moments required for the calculation are available since

$$\begin{aligned}\boldsymbol{\Phi}_{lr}\boldsymbol{w}_r^{i-1} &= \mathrm{E}\left[\boldsymbol{z}_l(d_r^{i-1})^*\right] \\ (\boldsymbol{w}_r^{i-1})^H\boldsymbol{\Phi}_{rr}\boldsymbol{w}_r^{i-1} &= \mathrm{E}\left[\|d_r^{i-1}\|^2\right].\end{aligned}$$

The expression $(\boldsymbol{w}_r^{i-1})^H\tilde{\boldsymbol{h}}_r^{\,l} = \frac{(\boldsymbol{w}_r^{i-1})^H\boldsymbol{h}_r}{h_{l,1}}$ equals the RTF between the desired source components at d_r^{i-1} and $z_{l,1}$. It can be estimated by exploiting the non-stationarity of speech or by the GEVD. The minimization of the right Lagrangian is solved in a similar manner. The algorithm is summarized in Alg. 1, where we also define

$$\bar{\boldsymbol{\Phi}}_{rr}^{i} = \boldsymbol{\Phi}_{rr} - \frac{\boldsymbol{\Phi}_{lr}\boldsymbol{w}_l^{i}(\boldsymbol{w}_l^{i})^H\boldsymbol{\Phi}_{lr}}{(\boldsymbol{w}_l^{i})^H\boldsymbol{\Phi}_{ll}\boldsymbol{w}_l^{i}} \tag{3.9}$$

$$\bar{\boldsymbol{h}}_r^{r,i} = \frac{(\boldsymbol{w}_l^{i})^H\tilde{\boldsymbol{h}}_l^{\,r}\boldsymbol{\Phi}_{rl}\boldsymbol{w}_l^{i}}{(\boldsymbol{w}_l^{i})^H\boldsymbol{\Phi}_{ll}\boldsymbol{w}_l^{i}} - \tilde{\boldsymbol{h}}_r^{\,r}. \tag{3.10}$$

The algorithm is initialized with the monaural MVDR on the left side. In the following section the proposed algorithm is proved to converge to the binaural MVDR beamformer in the rank-1 case.

3.1.4 Convergence of the distributed MVDR to the binaural MVDR

The variance of the MVDR output is denoted by $\mathcal{J}(\boldsymbol{w}) = \boldsymbol{w}^H\boldsymbol{\Phi}_{zz}\boldsymbol{w}$. Consider the ith iteration at the left side $\mathcal{J}\left(\boldsymbol{w}^{l,i}\right)$, where $\boldsymbol{w}^{l,i} = \begin{bmatrix} (\boldsymbol{w}_l^i)^T & \alpha i(\boldsymbol{w}_r^{i-1})^T \end{bmatrix}^T$. By manipulation of the left constraint $(\boldsymbol{w}^{l,i})^H\tilde{\boldsymbol{h}}^{\,l} = 1$ we obtain that $\left((\tilde{h}_{l,1}^r)^*\boldsymbol{w}^{l,i}\right)^H\tilde{\boldsymbol{h}}^{\,r} = 1$. Since $(\tilde{h}_{l,1}^r)^*\boldsymbol{w}^{l,i}$ belongs to the minimization range of the right side, the MVDR variance after updating the right weight coefficients (at the ith iteration) is upper bounded by $\mathcal{J}\left(\boldsymbol{w}^{r,i}\right) \leq \mathcal{J}\left((\tilde{h}_{l,1}^r)^*\boldsymbol{w}^{l,i}\right) = |(\tilde{h}_{l,1}^r)|^2\mathcal{J}\left(\boldsymbol{w}^{l,i}\right)$, where $\boldsymbol{w}^{r,i} = \begin{bmatrix} \alpha_r^i(\boldsymbol{w}_l^i)^T & (\boldsymbol{w}_r^i)^T \end{bmatrix}^T$. By manipulation of the right con-

straint $(\boldsymbol{w}^{r,i})^H\tilde{\boldsymbol{h}}^r = 1$ we obtain that $\left((\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,i}\right)^H \tilde{\boldsymbol{h}}^l = 1$. Therefore the MVDR variance after updating the left weight coefficients (at iteration $i+1$) is upper bounded by $\mathcal{J}\left(\boldsymbol{w}^{l,i+1}\right) \leq \mathcal{J}\left((\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,i}\right) = |(\tilde{h}^l_{r,1})|^2\mathcal{J}\left(\boldsymbol{w}^{r,i}\right)$. We therefore obtain the following inequality $\mathcal{J}\left(\boldsymbol{w}^{l,i+1}\right) \leq |(\tilde{h}^l_{r,1})|^2|(\tilde{h}^r_{l,1})|^2\mathcal{J}\left(\boldsymbol{w}^{l,i+1}\right)$, and since $\tilde{h}^l_{r,1} = (\tilde{h}^r_{l,1})^{-1}$ we conclude that the variance of the left MVDR is monotonically non-increasing $\mathcal{J}\left(\boldsymbol{w}^{l,i+1}\right) \leq \mathcal{J}\left(\boldsymbol{w}^{l,i}\right)$. In a similar way, we conclude that the variance of the right MVDR is also monotonically non-increasing $\mathcal{J}\left(\boldsymbol{w}^{r,i+1}\right) \leq \mathcal{J}\left(\boldsymbol{w}^{r,i}\right)$. $\mathcal{J}(\boldsymbol{w})$ is trivially lower bounded by 0 and therefore $\mathcal{J}(\boldsymbol{w}^{l,\infty})$, $\mathcal{J}(\boldsymbol{w}^{r,\infty})$ must converge. Consider the above inequalities for $i \to \infty$:

$$\mathcal{J}\left(\boldsymbol{w}^{l,\infty}\right) \leq \mathcal{J}\left((\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,\infty}\right) \tag{3.11a}$$

$$\mathcal{J}\left(\boldsymbol{w}^{r,\infty}\right) \leq \mathcal{J}\left((\tilde{h}^r_{l,1})^*\boldsymbol{w}^{l,\infty}\right). \tag{3.11b}$$

Dividing Eq. (3.11b) by $|\tilde{h}^r_{l,1}|^2$, noting that $\tilde{h}^r_{l,1}\tilde{h}^l_{r,1} = 1$ and combining both inequalities we have that $\mathcal{J}\left((\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,\infty}\right) \leq \mathcal{J}\left(\boldsymbol{w}^{l,\infty}\right) \leq \mathcal{J}\left((\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,\infty}\right)$ and due to the Squeeze theorem an equality holds

$$\mathcal{J}\left(\boldsymbol{w}^{l,\infty}\right) = \mathcal{J}\left((\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,\infty}\right). \tag{3.12}$$

Since $\mathcal{L}_l(\boldsymbol{w}^\infty_l, \alpha^\infty_l, \lambda^\infty_l)$ is the only local minimum as shown in Sec. 3.1.3, and since $(\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,\infty}$ belongs to the minimization range they coincide $\boldsymbol{w}^{l,\infty} = (\tilde{h}^l_{r,1})^*\boldsymbol{w}^{r,\infty}$, and the spatial filters after convergence are parallel. Notice that $\alpha^\infty_l = (\tilde{h}^l_{r,1})^*$ and $\alpha^\infty_r = (\tilde{h}^r_{l,1})^*$.

In order to prove that $\boldsymbol{w}^{ol} = \boldsymbol{w}^{l,\infty}$ we define the projection matrix to the desired signal subspace $\boldsymbol{P}^{\parallel} = \frac{\boldsymbol{h}\boldsymbol{h}^H}{\boldsymbol{h}^H\boldsymbol{h}}$ and to its null subspace $\boldsymbol{P}^{\perp} = \boldsymbol{I} - \frac{\boldsymbol{h}\boldsymbol{h}^H}{\boldsymbol{h}^H\boldsymbol{h}}$, where $\boldsymbol{I}$ is the identity matrix. Notice that the left and right constraints guarantee that at any iteration $i > 1$ the parallel components remain constant $\boldsymbol{P}^{\parallel}\boldsymbol{w}^{l,i} = \boldsymbol{w}^{l\parallel} = \boldsymbol{w}^{ol\parallel}$, $\boldsymbol{P}^{\parallel}\boldsymbol{w}^{r,i} = \boldsymbol{w}^{r\parallel} = \boldsymbol{w}^{or\parallel}$. Substitute $\boldsymbol{w}^{l,i} = (\boldsymbol{P}^{\parallel} + \boldsymbol{P}^{\perp})\boldsymbol{w}^{l,i} = \boldsymbol{w}^{l\parallel} + \boldsymbol{w}^{l\perp,i}$ in the left Lagrangian (3.6)

$$\begin{aligned}\mathcal{L}_l(\boldsymbol{w}^i_l, \alpha^i_l, \lambda^i_l) &= (\boldsymbol{w}^{l\parallel} + \boldsymbol{w}^{l\perp,i})^H\boldsymbol{\Phi}_{zz}(\boldsymbol{w}^{l\parallel} + \boldsymbol{w}^{l\perp,i}) \\ &+ \lambda^i_l\left((\boldsymbol{w}^{l\parallel})^H\tilde{\boldsymbol{h}}^l_l + (\boldsymbol{w}^{r\parallel})^H\tilde{\boldsymbol{h}}^l_r - 1\right) \\ &+ (\lambda^i_l)^*\left((\tilde{\boldsymbol{h}}^l_l)^H\boldsymbol{w}^{l\parallel}_l + (\tilde{\boldsymbol{h}}^l_r)^H\boldsymbol{w}^{l\parallel}_r - 1\right).\end{aligned} \tag{3.13}$$

The partial derivative $\frac{\partial \mathcal{L}_l}{\partial \boldsymbol{w}_l^{l\perp,i}}$ equals:

$$\frac{\partial \mathcal{L}_l}{\partial \boldsymbol{w}_l^{l\perp,i}} = \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\parallel}\right)_{ll}\boldsymbol{w}_l^{l\parallel} + \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\perp}\right)_{ll}\boldsymbol{w}_l^{l\perp,i} + \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\parallel}\right)_{lr}\boldsymbol{w}_r^{l\parallel} + \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\perp}\right)_{ll}\boldsymbol{w}_r^{l\perp,i}. \tag{3.14}$$

In a similar manner we reformulate $\mathcal{L}_r(\boldsymbol{w}_r^i, \alpha_r^i, \lambda_r^i)$ and evaluate its partial derivative $\frac{\partial \mathcal{L}_r}{\partial \boldsymbol{w}_r^{l\perp,i}}$

$$\frac{\partial \mathcal{L}_r}{\partial \boldsymbol{w}_r^{l\perp,i}} = \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\parallel}\right)_{ll}\boldsymbol{w}_l^{r\parallel} + \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\perp}\right)_{ll}\boldsymbol{w}_r^{l\perp,i} + \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\parallel}\right)_{lr}\boldsymbol{w}_r^{r\parallel} + \left(\boldsymbol{P}^{\perp}\boldsymbol{\Phi}_{zz}\boldsymbol{P}^{\perp}\right)_{ll}\boldsymbol{w}_r^{r\perp,i}. \tag{3.15}$$

Hence, the partial derivatives with respect to the left part of the orthogonal weights simultaneously equal zero $\frac{\partial \mathcal{L}_r}{\partial \boldsymbol{w}_r^{l\perp,i}}|\boldsymbol{w}^{r,\infty} = \alpha_r^{\infty}\frac{\partial \mathcal{L}_l}{\partial \boldsymbol{w}_l^{l\perp,i}}|\boldsymbol{w}^{l,\infty} = 0$ when $i \to \infty$. The right partial derivatives simultaneously equal zero in a similar manner. Finally, the global minimum is reached, since the parallel part of the weights equals its optimum, and since all the partial derivatives according to the left and right orthogonal parts of the weights equal zero. Furthermore, since the global Lagrangian under minimization has a single minimum, it has been reached. Hence, it is concluded that

$$\boldsymbol{w}^{l,\infty} = \boldsymbol{w}^{ol} \tag{3.16a}$$

$$\boldsymbol{w}^{r,\infty} = \boldsymbol{w}^{or}\blacksquare \tag{3.16b}$$

3.1.5 Experimental Study

The proposed algorithm was evaluated using both narrow-band stationary signals and speech signals in a reverberant room. The proposed method was compared with the closed-form binaural MVDR, and the closed-form monaural MVDR. The narrow-band scenario is comprised of one desired source and twointerfering sources received by two sub-arrays each comprised of 2 sensors with random ATFs. The sources are generated directly at the STFT domain as complex Gaussian random variables uncorrelated between time frames. A spatially white sensor noise is added to the received signals. The SIR and SNR were set to 0dB and 20dB, respectively. The various correlation matrices were assumed to be known. In Fig. 3.2(a) a comparison between the noise variance of the proposed distributed MVDR and the monaural

Algorithm 1 Distributed binaural MVDR

begin

  for $i = 1, 2, ..$ **do**

    if $i = 1$ **then**

      $\boldsymbol{w}_l^1 = \boldsymbol{\Phi}_{ll}^{-1}\tilde{\boldsymbol{h}}_l^{\,l}\left((\tilde{\boldsymbol{h}}_l^{\,l})^H\boldsymbol{\Phi}_{ll}^{-1}\tilde{\boldsymbol{h}}_l^{\,l}\right)^{-1}$

      $\alpha_l^1 = 0$

      $d_r^1 = \mathbf{0}$

    else

      $\lambda_l^i = \left(-\|\bar{\boldsymbol{h}}_l^{l,i}\|^2_{(\bar{\boldsymbol{\Phi}}_{ll}^{\,i})^{-1}} - \frac{\|(\tilde{\boldsymbol{h}}_r^{\,l})^H\boldsymbol{w}_r^{i-1}\|^2}{(\boldsymbol{w}_r^{i-1})^H\boldsymbol{\Phi}_{rr}\boldsymbol{w}_r^{i-1}}\right)^{-1}$

      $\boldsymbol{w}_l^i = \lambda_l^i(\bar{\boldsymbol{\Phi}}_{ll}^{\,i})^{-1}\bar{\boldsymbol{h}}_l^{l,i}$

      $\alpha_l^i = -\frac{(\boldsymbol{w}_r^{i-1})\boldsymbol{\Phi}_{rl}\boldsymbol{w}_l^i + \lambda_l^i(\boldsymbol{w}_r^{i-1})^H\tilde{\boldsymbol{h}}_r^{\,l}}{(\boldsymbol{w}_r^{i-1})^H\boldsymbol{\Phi}_{rr}\boldsymbol{w}_r^{i-1}}$

      $\boldsymbol{w}_l^i = \frac{\boldsymbol{w}_l^i}{(\tilde{\boldsymbol{h}}_l^{\,l})^H\boldsymbol{w}_l^i}$

    end

    $d_l^i = (\boldsymbol{w}_l^i)^H\boldsymbol{z}_l$

    $y_l^i = (\boldsymbol{w}_l^i)^H\boldsymbol{z}_l + (\alpha_l^i)^* d_r^{i-1}$

    Transmit d_l^i to the right side

    $\lambda_r^i = \left(-\|\bar{\boldsymbol{h}}_r^{r,i}\|^2_{(\bar{\boldsymbol{\Phi}}_{rr}^{\,i})^{-1}} - \frac{\|(\tilde{\boldsymbol{h}}_l^{\,r})^H\boldsymbol{w}_l^i\|^2}{(\boldsymbol{w}_l^i)^H\boldsymbol{\Phi}_{ll}\boldsymbol{w}_l^i}\right)^{-1}$

    $\boldsymbol{w}_r^i = \lambda_r^i(\bar{\boldsymbol{\Phi}}_{rr}^{\,i})^{-1}\bar{\boldsymbol{h}}_r^{r,i}$

    $\alpha_r^i = -\frac{(\boldsymbol{w}_l^i)\boldsymbol{\Phi}_{lr}\boldsymbol{w}_r^i + \lambda_r^i(\boldsymbol{w}_l^i)^H\tilde{\boldsymbol{h}}_l^{\,r}}{(\boldsymbol{w}_l^i)^H\boldsymbol{\Phi}_{ll}\boldsymbol{w}_l^i}$

    $\boldsymbol{w}_r^i = \frac{\boldsymbol{w}_r^i}{(\tilde{\boldsymbol{h}}_r^{\,r})^H\boldsymbol{w}_r^i}$

    $d_r^i = (\boldsymbol{w}_r^i)^H\boldsymbol{z}_r$

    $y_r^i = (\boldsymbol{w}_r^i)^H\boldsymbol{z}_r + (\alpha_r^i)^* d_l^i$

    Transmit d_r^i to the left side

  end

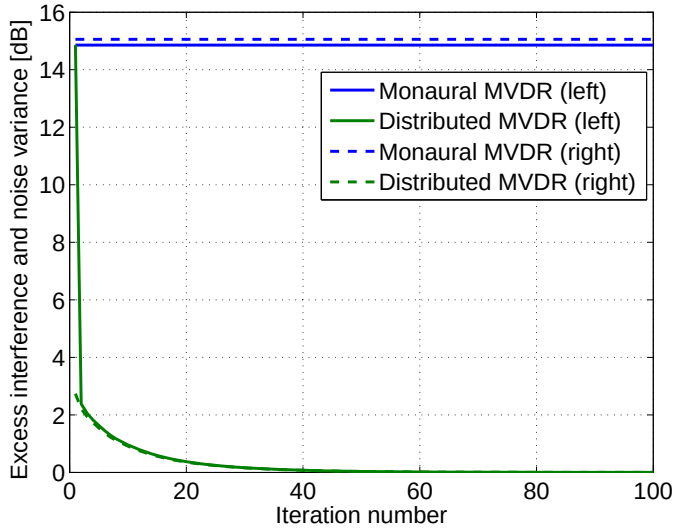
end

MVDR normalized by the noise variance of the binaural MVDR is shown for the left and right laterals. 50 Monte-carlo trials were used to produce the graphs. The noise variance is plotted versus the iteration number. It is clear from the figure that the proposed algorithm converges to the binaural MVDR.

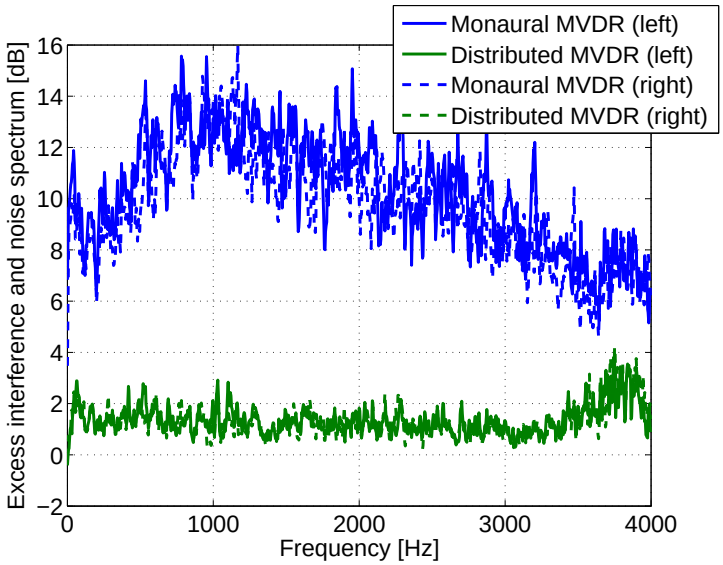
The wide-band speech scenario is comprised of one desired speaker and two interference signals received by two sub-arrays comprised of 2 microphones in a simulated $4\times4\times3\text{m}^3$ room environment, with a reverberation time set to 150ms. The SIR and SNR were set to 15dB and 60dB, respectively. The two sub-arrays were oriented in parallel. The inter sub-array

distance was set to 17cm. The distance between the 2 microphones at each sub-array was set to 5cm. The signals were sampled at 8kHz and were transformed into the STFT domain with 2048 points and 75% overlap. The various algorithms operated in each frequency bin independently. The signals were filtered in the time-domain[1]. The noise and speech correlation matrices were estimated in noise-only and in speech and noise segments, respectively. A simple energy threshold voice activity detector (VAD) was utilized. The desired signal RTF was estimated by selecting the major eigenvector of the GEVD of the received signals correlation matrix and the noise-only correlation matrix. The average noise PSD at the outputs of the proposed distributed algorithm and the monaural MVDR algorithms normalized by the noise PSD at the output of the binaural MVDR is depicted in Fig. 3.2(b). The graphs were obtained by averaging 20 Monte-Carlo experiments. The number of iterations of the distributed MVDR algorithm was set to 10. Due to estimation errors and filter windowing the noise PSD of the proposed distributed MVDR is 1.5dB higher than the noise PSD of the binaural MVDR. In Fig. 3.3 the distributed MVDR is compared with the binaural and monaural MVDR by subjective assessment of speech sonograms. It is clearly seen that the proposed distributed binaural MVDR outperforms the monaural MVDR, and that its performance is equivalent to the binaural MVDR.

[1]The noise PSD and the ATFs were assumed to be time-invariant, hence the MVDR beamformers were also time-invariant.

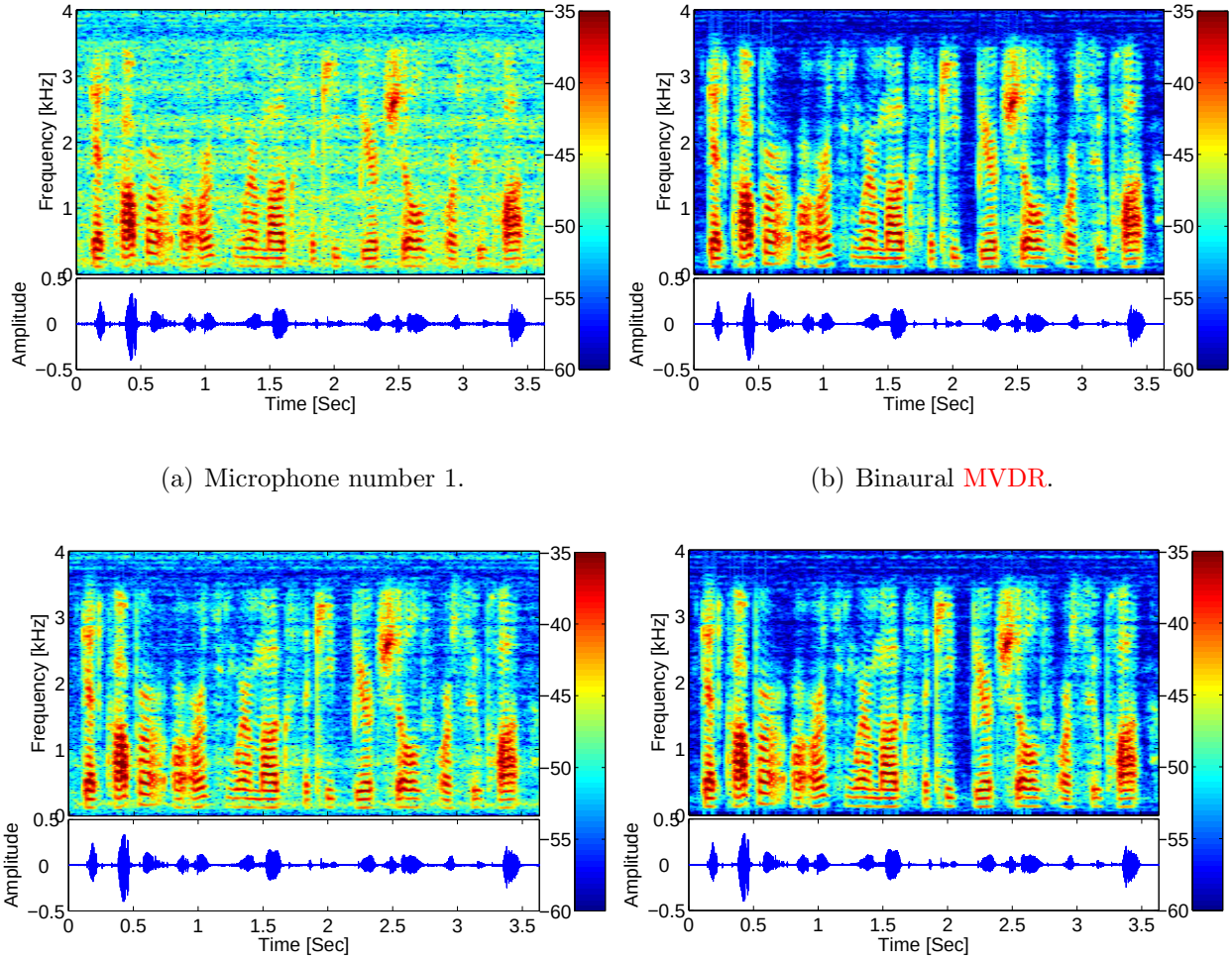


(a) Noise variance normalized by the noise variance at the output of the binaural MVDR.



(b) Noise PSD normalized by the noise PSD at the output of the binaural MVDR.

Figure 3.2: Excess noise variance in narrow-band scenario and excess noise PSD in wide-band scenario.



(a) Microphone number 1.

(b) Binaural MVDR.

(c) Monaural MVDR.

(d) Distributed MVDR.

Figure 3.3: Sonograms of left signals in an example scenario.

3.1.6 Conclusions

A novel distributed MVDR BF has been introduced. The proposed method reduces the energy consumption by requiring a single transmission channel between two sub-arrays, regardless of the number of microphones. The algorithm is proved to converge to the optimal binaural MVDR when the desired source correlation matrix has a rank-1. The algorithm is applied to the binaural hearing aid problem. The experimental study demonstrates the superior performance of the proposed algorithm in comparison with the monaural MVDR. The convergence to the binaural MVDR is analytically proven for the narrow-band case. A time recursive version of the algorithm can be obtained by using a recursive estimation of

the involved correlation matrices.

3.2 Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks

In this section we propose a distributed multiple constraints GSC for speech enhancement in an N-node fully connected WASN comprising $\bar{M}$ microphones. Our algorithm is designed to operate in reverberant environments with P constrained speakers (including both desired and competing speakers). Rather than broadcasting $\bar{M}$ microphone signals, a significant communication bandwidth reduction is obtained by performing *local* beamforming at the nodes, and utilizing only $N + P$ transmission channels. Each node processes its own microphone signals together with the transmitted signals. The GSC-form implementation, by separating the constraints and the minimization, enables the adaptation of the BF during speech-absent time segments, and relaxes the requirement of other distributed LCMV based algorithms to re-estimate the sources' RTFs after each iteration. We provide a full convergence proof of the proposed structure to the centralized GSC-BF. An extensive experimental study of both narrowband and (wideband) speech signals verifies the theoretical analysis.

The structure of this section is as follows. In Sec. 3.2.1, the problem is formulated. In Sec. 3.2.2, a closed-form and a GSC structure of the centralized LCMV-BF are presented. We show that, under certain conditions, an LCMV which operates on a transformation of the inputs is equivalent to the regular BF. In Sec. 3.2.3, we derive the DGSC algorithm. The latter is based on a specific transformation which allows to reformulate the centralized BF as a sum of local GSC-BFs. The proposed algorithm makes use of shared signals, one for each source, which are broadcast in the WASN. We give an analytical proof of the equivalence between the DGSC and the centralized GSC-BF. In Sec. 3.2.4, we propose a scheme for constructing the shared signals. We compare the proposed DGSC and the LC-DANSE in Sec. 3.2.5. An extensive experimental study, which verifies the equivalence of the DGSC and the centralized GSC, is presented in Sec. 3.2.6. Conclusions are drawn in Sec. 3.2.7.

3.2.1 Problem formulation

Consider a WASN of microphones comprised of N nodes. Denote the number of microphones in the nth node by $\bar{M}_n$. The total number of microphones is denoted $\bar{M}$ and equals

$$\bar{M} \triangleq \sum_{n=1}^{N} \bar{M}_n. \tag{3.17}$$

The problem is formulated in the STFT domain where k denotes the frequency index and ℓ denotes the time-frame index. The vector of signals received by the microphones of all nodes is $\bar{\boldsymbol{z}}(\ell,k)$. It is composed by concatenating the microphone signals of all nodes:

$$\bar{\boldsymbol{z}}(\ell,k) \triangleq \left[\begin{array}{ccc} \bar{\boldsymbol{z}}_1^T(\ell,k) & \cdots & \bar{\boldsymbol{z}}_N^T(\ell,k) \end{array} \right]^T \tag{3.18}$$

where $\bar{\boldsymbol{z}}_n(\ell,k)$ is an $\bar{M}_n \times 1$ vector consisting of locally received signals at the nth node. The vector of all received signals is given by:

$$\bar{\boldsymbol{z}}(\ell,k) \triangleq \bar{\boldsymbol{H}}(\ell,k)\boldsymbol{s}(\ell,k) + \bar{\boldsymbol{v}}(\ell,k) \tag{3.19}$$

where

$$\boldsymbol{s}(\ell,k) \triangleq \left[\begin{array}{ccc} s^1(\ell,k) & \cdots & s^P(\ell,k) \end{array} \right]^T \tag{3.20}$$

is a $P \times 1$ vector comprised of the speech sources, and

$$\bar{\boldsymbol{H}}(\ell,k) \triangleq \left[\begin{array}{ccc} \bar{\boldsymbol{h}}^1(\ell,k) & \cdots & \bar{\boldsymbol{h}}^P(\ell,k) \end{array} \right] \tag{3.21}$$

is an $\bar{M} \times P$ matrix which columns are the ATFs relating the P speakers and the $\bar{M}$ microphones. The vector $\bar{\boldsymbol{v}}(\ell,k)$ is a vector of interfering signals picked up by the microphones. Assuming that the P speakers' signals and the noise sources are uncorrelated, the $\bar{M} \times \bar{M}$ dimensional covariance matrix of the received signals may be written as:

$$\bar{\boldsymbol{\Phi}}_{zz}(\ell,k) \triangleq \bar{\boldsymbol{H}}(\ell,k)\boldsymbol{\Gamma}(\ell,k)\bar{\boldsymbol{H}}^H(\ell,k) + \bar{\boldsymbol{\Phi}}_{vv}(k,\ell) \tag{3.22}$$

where $(\bullet)^H$ denotes the conjugate-transpose operator, $\boldsymbol{\Gamma}(\ell,k) = \text{diag}\left\{\lambda^1(\ell,k),\ldots,\lambda^P(\ell,k)\right\}$ is the $P \times P$ dimensional covariance matrix of the P speech signals and $\bar{\boldsymbol{\Phi}}_{vv}(\ell,k)$ is the covariance matrix of the noise. Note that multiple speakers and noise sources may be si-

multaneously active at each frequency bin. We assume that the network is fully connected, hence any transmitted signal is available to all nodes. In cases that the network is not fully connected a hierarchial algorithm, for example based on a spanning tree of the network, can be sought. However, this is beyond the scope of the current contribution. As an example for a distributed algorithm in a partially connected WASN please refer to [104]. The locations of the speakers are assumed static, therefore their corresponding ATFs are time-invariant, and hence the frame index is omitted in $\bar{\boldsymbol{H}}(k)$. The algorithm is applied to each frequency bin independently. For brevity, the index k is hereafter omitted. The noise statistics is assumed to vary significantly slower than the convergence-time of the algorithm. For brevity, the index ℓ is also omitted from $\bar{\boldsymbol{\Phi}}_{vv}$ hereafter.

Denote the set of microphone indexes at the nth node by $\bar{\mathcal{M}}_n \triangleq \{m_n(1), \ldots, m_n(\bar{M}_n)\}$, where $\bar{M}_n \triangleq |\bar{\mathcal{M}}_n|$ and $|\bullet|$ denotes the number of elements in a set. The vector of the received signals at the nth node is given by

$$\bar{\boldsymbol{z}}_n(\ell) = \boldsymbol{T}_n^H \bar{\boldsymbol{z}}(\ell) \tag{3.23}$$

where $\boldsymbol{T}_n$ is an $\bar{M} \times \bar{M}_n$ selection matrix which extracts the $\bar{M}_n$ entries that correspond to the microphone indexes of the nth node:

$$\boldsymbol{T}_n = \left[\begin{array}{c|c|c} \boldsymbol{0}_{\bar{M}_n \times (\sum_{n'=1}^{n-1} \bar{M}_{n'})} & \boldsymbol{I}_{\bar{M}_n} & \boldsymbol{0}_{\bar{M}_n \times (\sum_{n'=n+1}^{N} \bar{M}_{n'})} \end{array} \right]^T \tag{3.24}$$

and $\boldsymbol{I}_m$ is an $m \times m$ identity matrix.

3.2.2 An equivalent centralized LCMV-BF

In the following, the centralized LCMV-BF is formulated. We show that under certain conditions, an LCMV-BF which operates on a transformation of the inputs is equivalent to the LCMV-BF which directly processes the microphone signals. A common design relaxation of using the RTFs rather than the ATFs is formulated, and the GSC-form implementation is defined. The distributed algorithm, derived in Sec. 3.2.3, will be based on a specific transformation matrix, that will conveniently split the centralized BF into a sum of N BFs. Each of the BFs utilizes only local microphones and P shared signals, generated as a linear combination of the local microphone signals in some remote nodes. Together with the transmission of the N local BF outputs, a total of $N + P$ transmission channels is required.

The centralized LCMV-BF, denoted $\bar{\boldsymbol{w}}_{\text{LCMV}}$, is given by:

$$\bar{\boldsymbol{w}}_{\text{LCMV}} \triangleq \underset{\left\{\boldsymbol{w};\bar{\boldsymbol{H}}^H\boldsymbol{w}=\boldsymbol{g}\right\}}{\text{argmin}} \left\{\boldsymbol{w}^H\bar{\boldsymbol{\Phi}}_{vv}\boldsymbol{w}\right\} \tag{3.25}$$

where the global constraints set is

$$\bar{\boldsymbol{H}}^H\bar{\boldsymbol{w}} = \boldsymbol{g} \tag{3.26}$$

and $\boldsymbol{g}$ is a $P\times 1$ desired response vector. Typically, the desired response vector $\boldsymbol{g}$ is comprised of values of zeros and ones, where a value of 1 is associated with a desired speaker and a value of 0 is associated with an interfering speaker. In this case the BF is required to yield a combination of all the desired speakers while mitigating the interfering speakers and the noise. Generally, $\boldsymbol{g}$ can be any arbitrary $P \times 1$ vector. We assume that the ATFs are linearly independent, i.e., the column rank of the constraints matrix $\bar{\boldsymbol{H}}$ is P. In practice, when $\bar{M} \gg P$ the latter assumption usually holds, however, of course it is not guaranteed. In cases for which the ATFs are linearly dependent, the constraints set might consist of contradicting requirements. Hence, no solution that satisfies all constraints can be obtained. When contradicting constraints exist, the system designer has to compromise and alleviate the contradiction by reducing the number of constraints. The closed-form solution of (3.25) is given by Van Veen and Buckley in [25]:

$$\bar{\boldsymbol{w}}_{\text{LCMV}} = \bar{\boldsymbol{\Phi}}_{vv}^{-1}\bar{\boldsymbol{H}}\left(\bar{\boldsymbol{H}}^H\bar{\boldsymbol{\Phi}}_{vv}^{-1}\bar{\boldsymbol{H}}\right)^{-1}\boldsymbol{g} \tag{3.27}$$

where we assume that $\bar{\boldsymbol{\Phi}}_{vv}$ is invertible since one of its components is a spatially white sensor noise.

The output of the LCMV-BF is given by:

$$\bar{y}_{\text{LCMV}}(\ell) = \sum_{p=1}^{P}(g^p)^* s^p(\ell) + \bar{\boldsymbol{w}}_{\text{LCMV}}^H\bar{\boldsymbol{v}}(\ell) \tag{3.28}$$

where $\boldsymbol{g} = \begin{bmatrix} g^1 & \cdots & g^P \end{bmatrix}^T$. Note that the output comprises of the sum of the constrained sources weighted by their corresponding desired responses and a residual noise component.

Suppose that rather than $\bar{\boldsymbol{z}}(\ell)$, a linear transformation of the inputs is available:

$$\boldsymbol{z}(\ell) \triangleq \boldsymbol{U}^H \bar{\boldsymbol{z}}(\ell) \tag{3.29}$$

where $\boldsymbol{U}^H$ is an $M \times \bar{M}$ matrix and $M > \bar{M}$. Assuming that the column-subspace of $\boldsymbol{U}^H$ is full rank, i.e., its rank is $\bar{M}$, we will show that the LCMV-BFs in the original and in the transformed domains are equivalent. Denote the following terms in the transformed domain:

$$\tilde{\boldsymbol{H}} \triangleq \boldsymbol{U}^H \bar{\boldsymbol{H}} \tag{3.30a}$$

$$\boldsymbol{\Phi}_{vv} \triangleq \boldsymbol{U}^H \bar{\boldsymbol{\Phi}}_{vv} \boldsymbol{U}. \tag{3.30b}$$

Consider the following constraints set in the transformed domain:

$$\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{w}} = \boldsymbol{g}. \tag{3.31}$$

According to the fundamental theorem of linear algebra, any $M \times 1$ BF, $\tilde{\boldsymbol{w}}$, in the transformed domain can be expressed as the sum of two components:

$$\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{w}}^u + \tilde{\boldsymbol{w}}^{uc} \tag{3.32}$$

where $\tilde{\boldsymbol{w}}^u$ and $\tilde{\boldsymbol{w}}^{uc}$ lie in the column-subspace of $\boldsymbol{U}^H$ and its complementary subspace, respectively. Similarly to (3.25), the LCMV criterion in the transformed domain is:

$$\tilde{\boldsymbol{w}}_{\text{LCMV}} \triangleq \underset{\left\{\tilde{\boldsymbol{w}}; \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{w}} = \boldsymbol{g}\right\}}{\operatorname{argmin}} \left\{\tilde{\boldsymbol{w}}^H \boldsymbol{\Phi}_{vv} \tilde{\boldsymbol{w}}\right\}. \tag{3.33}$$

Note that from the definition of $\tilde{\boldsymbol{H}}$ and $\boldsymbol{\Phi}_{vv}$ in (3.30a) and (3.30b), their columns lie in the column-subspace of $\boldsymbol{U}^H$. Hence, substituting (3.32) in the transformed constraint set (3.31) and in the minimization of the transformed LCMV-BF (3.33) yields:

$$\tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{w}}^u = \boldsymbol{g} \tag{3.34a}$$

$$\tilde{\boldsymbol{w}}^u_{\text{LCMV}} = \underset{\left\{\tilde{\boldsymbol{w}}^u; \tilde{\boldsymbol{H}}^H \tilde{\boldsymbol{w}}^u = \boldsymbol{g}\right\}}{\operatorname{argmin}} \left\{(\tilde{\boldsymbol{w}}^u)^H \boldsymbol{\Phi}_{vv} \tilde{\boldsymbol{w}}^u\right\} \tag{3.34b}$$

$$\tilde{\boldsymbol{w}}_{\text{LCMV}} = \tilde{\boldsymbol{w}}^{uc} + \tilde{\boldsymbol{w}}^u_{\text{LCMV}} \tag{3.34c}$$

where the orthogonal component $\tilde{\boldsymbol{w}}^{uc}$ can be chosen arbitrarily, since it affects neither the noise power at the output nor the satisfaction of the constraints set. Any $\tilde{\boldsymbol{w}}^{u}$ can be expressed as a linear combination of the columns of $\boldsymbol{U}^H$:

$$\tilde{\boldsymbol{w}}^{u} \triangleq \boldsymbol{U}^H \boldsymbol{\omega} \tag{3.35}$$

where $\boldsymbol{\omega}$ is an $\bar{M} \times 1$ vector.

Substituting (3.35) in (3.34a),(3.34b),(3.34c), $\tilde{\boldsymbol{w}}_{\text{LCMV}}$ becomes:

$$\begin{aligned}\tilde{\boldsymbol{w}}_{\text{LCMV}} = \tilde{\boldsymbol{w}}^{uc} \\ + \boldsymbol{U}^H \underset{\left\{\boldsymbol{\omega};\left(\boldsymbol{U}\tilde{\boldsymbol{H}}\right)^H \boldsymbol{\omega}=\boldsymbol{g}\right\}}{\text{argmin}} \left\{\boldsymbol{\omega}^H \boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H\boldsymbol{\omega}\right\}.\end{aligned} \tag{3.36}$$

Note that

$$\boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H = \boldsymbol{U}\boldsymbol{U}^H\bar{\boldsymbol{\Phi}}_{vv}\boldsymbol{U}\boldsymbol{U}^H \tag{3.37}$$

is a full-rank $\bar{M} \times \bar{M}$ matrix since both $\boldsymbol{U}\boldsymbol{U}^H$ and $\bar{\boldsymbol{\Phi}}_{vv}$ are $\bar{M} \times \bar{M}$ dimensional rank-$\bar{M}$ matrices. Hence, similarly to (3.27), the closed-form LCMV-BF of (3.36) in the transformed domain equals:

$$\begin{aligned}\tilde{\boldsymbol{w}}_{\text{LCMV}} =& \tilde{\boldsymbol{w}}^{uc} + \boldsymbol{U}^H \left(\boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H\right)^{-1} \\ & \times \boldsymbol{U}\tilde{\boldsymbol{H}} \left(\left(\boldsymbol{U}\tilde{\boldsymbol{H}}\right)^H \left(\boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H\right)^{-1} \boldsymbol{U}\tilde{\boldsymbol{H}}\right)^{-1} \boldsymbol{g}.\end{aligned} \tag{3.38}$$

Substituting (3.30a),(3.30b) and (3.27) in (3.38) yields

$$\tilde{\boldsymbol{w}}_{\text{LCMV}} = \boldsymbol{U}^H \left(\boldsymbol{U}\boldsymbol{U}^H\right)^{-1} \bar{\boldsymbol{w}}_{\text{LCMV}} + \tilde{\boldsymbol{w}}^{uc}, \tag{3.39}$$

where we also used the invertibility of $\boldsymbol{U}\boldsymbol{U}^H$. It can be easily deduced that the BFs in the original and transformed domains are equivalent as their outputs coincide:

$$\tilde{\boldsymbol{w}}^H_{\text{LCMV}} \boldsymbol{z}(\ell) = \bar{\boldsymbol{w}}^H_{\text{LCMV}} \bar{\boldsymbol{z}}(\ell). \tag{3.40}$$

In practice the ATFs of the speakers are unknown, and are difficult to estimate. A practical solution can be obtained by replacing the sources in (3.28) with filtered versions

thereof [39, 51, 105, 95]. Let $h_{\text{ref}}^p;\ p=1,\ldots,P$ be such filters. The RTF of the pth source in the transformed domain is defined as:

$$\boldsymbol{h}^p \triangleq \frac{\tilde{\boldsymbol{h}}^p}{h_{\text{ref}}^p}. \tag{3.41}$$

The filters $h_{\text{ref}}^p;\ p=1,\ldots,P$ will be determined in Sec. 3.2.3. Note that these procedures may require non-overlapping activity patterns of the speakers.

Define the transformed ATF and RTF matrices of dimensions $M \times P$, respectively:

$$\tilde{\boldsymbol{H}} \triangleq \left[\begin{array}{ccc} \tilde{\boldsymbol{h}}^1 & \cdots & \tilde{\boldsymbol{h}}^P \end{array}\right] \tag{3.42a}$$

$$\boldsymbol{H} \triangleq \left[\begin{array}{ccc} \boldsymbol{h}^1 & \cdots & \boldsymbol{h}^P \end{array}\right]. \tag{3.42b}$$

The modified constraints set is finally given by substituting $\tilde{\boldsymbol{H}}$ by $\boldsymbol{H}$ in (3.26):

$$\boldsymbol{H}^H \boldsymbol{w}_{\text{LCMV}} = \boldsymbol{g}. \tag{3.43}$$

The modified centralized LCMV-BF (in the transformed domain), which satisfies the modified constraints set in (3.43), is denoted by $\boldsymbol{w}_{\text{LCMV}}$ and is given in closed-form, similarly to (3.38):

$$\begin{aligned} \boldsymbol{w}_{\text{LCMV}} =& \boldsymbol{w}^{uc} + \boldsymbol{U}^H \left(\boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H\right)^{-1} \\ & \times \boldsymbol{U}\boldsymbol{H}\left(\left(\boldsymbol{U}\boldsymbol{H}\right)^H \left(\boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H\right)^{-1} \boldsymbol{U}\boldsymbol{H}\right)^{-1} \boldsymbol{g}. \end{aligned} \tag{3.44}$$

where $\boldsymbol{w}^{uc}$ is an arbitrary vector lying in the null-subspace of the column-subspace of $\boldsymbol{U}^H$. Similarly to (3.34b),(3.34c) we identify that the component of $\boldsymbol{w}_{\text{LCMV}}$ which lies in the column-subspace of $\boldsymbol{U}^H$ is:

$$\begin{aligned} \boldsymbol{w}_{\text{LCMV}}^u =& \boldsymbol{U}^H \left(\boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H\right)^{-1} \\ & \times \boldsymbol{U}\boldsymbol{H}\left(\left(\boldsymbol{U}\boldsymbol{H}\right)^H \left(\boldsymbol{U}\boldsymbol{\Phi}_{vv}\boldsymbol{U}^H\right)^{-1} \boldsymbol{U}\boldsymbol{H}\right)^{-1} \boldsymbol{g}. \end{aligned} \tag{3.45}$$

The GSC-form implementation of (3.45), denoted centralized GSC-BF [35, 39], is obtained by splitting $\boldsymbol{w}^u$ into two components:

$$\boldsymbol{w}_{\text{LCMV}}^u \triangleq \boldsymbol{q}_{\text{GSC}} - \boldsymbol{B}_{\text{GSC}}\boldsymbol{f}_{\text{GSC}}. \tag{3.46}$$

Both $\boldsymbol{q}_{\text{GSC}}$ and the columns of $\boldsymbol{B}_{\text{GSC}}$ lie in the column-subspace of $\boldsymbol{U}^H$. The vector $\boldsymbol{q}_{\text{GSC}}$, denoted FBF, lies in the column-subspace of $\boldsymbol{H}$. $\boldsymbol{q}_{\text{GSC}}$ is responsible for maintaining the modified constraints set (3.43), and equals:

$$\boldsymbol{q}_{\text{GSC}} = \boldsymbol{H}\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{g}. \tag{3.47}$$

The BM, $\boldsymbol{B}_{\text{GSC}}$, blocks the RTFs of the constrained speakers. Explicitly,

$$\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{H} = \boldsymbol{0}. \tag{3.48}$$

Since the ranks of $\boldsymbol{U}^H$ and $\boldsymbol{H}$ are $\bar{M}$ and P, respectively, the rank of $\boldsymbol{B}_{\text{GSC}}$ is $\bar{M}-P$ and its dimensions are $M\times\left(\bar{M}-P\right)$. The BM is not unique and can be obtained in several ways, for example, as suggested in [25, 101], by applying the SVD. To construct the BM, the SVD is applied to the $\bar{M}\times P$ matrix $\boldsymbol{U}\boldsymbol{H}$, rather than $\boldsymbol{H}$, and then projected to the transformed domain. Using this procedure an $M\times(\bar{M}-P)$ BM is obtained. Denote the NC by an $(\bar{M}-P)\times 1$ vector $\boldsymbol{f}_{\text{GSC}}$. According to [25] it equals:

$$\boldsymbol{f}_{\text{GSC}} = \left(\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}_{\text{GSC}}\right)^{-1}\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{\Phi}_{vv}\boldsymbol{q}_{\text{GSC}}. \tag{3.49}$$

Note that the invertibility of $\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}_{\text{GSC}}$ is guaranteed by the definition (3.30b) and by the BM construction procedure above.

To enable the construction of the DGSC in Sec. 3.2.3, an extended GSC-structure is proposed:

$$\boldsymbol{w} \triangleq \boldsymbol{q} - \boldsymbol{B}\boldsymbol{f} \tag{3.50}$$

where the regular GSC components, $\boldsymbol{q}_{\text{GSC}}$, $\boldsymbol{B}_{\text{GSC}}$ and $\boldsymbol{f}_{\text{GSC}}$, are replaced by:

$$\boldsymbol{q} = \boldsymbol{q}_{\text{GSC}} + \boldsymbol{B}\boldsymbol{a} + \boldsymbol{q}^{uc} \tag{3.51a}$$

$$\boldsymbol{B} = \boldsymbol{B}_{\text{GSC}} + \boldsymbol{B}^{uc} \tag{3.51b}$$

$$\boldsymbol{f} = \left(\boldsymbol{B}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^H\boldsymbol{\Phi}_{vv}\boldsymbol{q}. \tag{3.51c}$$

Here, the regular FBF $\boldsymbol{q}_{\text{GSC}}$ is extended by the vectors $\boldsymbol{B}\boldsymbol{a}$ and $\boldsymbol{q}^{uc}$, and the regular BM $\boldsymbol{B}_{\text{GSC}}$ is extended by the matrix $\boldsymbol{B}^{uc}$. The extensions $\boldsymbol{q}^{uc}$ and $\boldsymbol{B}^{uc}$ lie in the columns null-subspace of the matrix $\boldsymbol{U}^H$, and $\boldsymbol{B}\boldsymbol{a}$ lies in null-subspace of $\boldsymbol{H}$. For any choice of $\boldsymbol{a}$,

$\boldsymbol{q}^{uc}$, $\boldsymbol{B}^{uc}$ the modified constraints set (3.43) is maintained. Note that the regular GSC can be obtained as a special case of (3.51a), (3.51b) and (3.51c) by setting $\boldsymbol{a}=\boldsymbol{0}$, $\boldsymbol{q}^{uc}=\boldsymbol{0}$ and $\boldsymbol{B}^{uc}=\boldsymbol{0}$. As will be seen in the sequel, the introduction of $\boldsymbol{a}\neq\boldsymbol{0}$, $\boldsymbol{q}^{uc}\neq\boldsymbol{0}$ and $\boldsymbol{B}^{uc}\neq\boldsymbol{0}$ will enable us to derive a distributed version of the GSC.

Now, we show that:

$$\boldsymbol{w}^H\boldsymbol{z}(\ell)=\boldsymbol{w}_{\text{LCMV}}^H\boldsymbol{z}(\ell) \tag{3.52}$$

i.e., that $\boldsymbol{w}$ and $\boldsymbol{w}_{\text{LCMV}}$ are equivalent. By substituting (3.51a), (3.51b), (3.51c) in (3.50), it is evident that:

$$\begin{aligned}\boldsymbol{w}=&\boldsymbol{q}_{\text{GSC}}+\boldsymbol{B}\boldsymbol{a}+\boldsymbol{q}^{uc}-\boldsymbol{B}\left(\boldsymbol{B}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^H\boldsymbol{\Phi}_{vv}\boldsymbol{q}\\ \overset{(3.49)}{=}&\boldsymbol{q}_{\text{GSC}}-\boldsymbol{B}_{\text{GSC}}\boldsymbol{f}_{\text{GSC}}\\ &+\boldsymbol{q}^{uc}-\boldsymbol{B}^{uc}\left(\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}_{\text{GSC}}\right)^{-1}\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{\Phi}_{vv}\boldsymbol{q}_{\text{GSC}}\\ \overset{(3.45)}{=}&\boldsymbol{w}_{\text{LCMV}}^u+\boldsymbol{w}^{uc}\end{aligned} \tag{3.53}$$

where $\boldsymbol{w}^{uc}$ is identified as:

$$\boldsymbol{w}^{uc}=\boldsymbol{q}^{uc}-\boldsymbol{B}^{uc}\left(\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}_{\text{GSC}}\right)^{-1}\boldsymbol{B}_{\text{GSC}}^H\boldsymbol{\Phi}_{vv}\boldsymbol{q}_{\text{GSC}}. \tag{3.54}$$

This concludes the proof of the equivalence between the extended and the regular GSC-structures.

The output signal of the proposed GSC-structure is given by:

$$\begin{aligned}y(\ell)\triangleq&\boldsymbol{w}^H\boldsymbol{z}(\ell)\\ =&y^{\text{FBF}}(\ell)-y^{\text{NC}}(\ell)\end{aligned} \tag{3.55}$$

where $y^{\text{FBF}}(\ell)$ and $y^{\text{NC}}(\ell)$ are the outputs of the upper and lower branches of the GSC, respectively:

$$y^{\text{FBF}}(\ell)\triangleq\boldsymbol{q}^H\boldsymbol{z}(\ell) \tag{3.56a}$$

$$y^{\text{NC}}(\ell)\triangleq\boldsymbol{f}^H\boldsymbol{u}(\ell) \tag{3.56b}$$

$$\boldsymbol{u}(\ell)\triangleq\boldsymbol{B}^H\boldsymbol{z}(\ell) \tag{3.56c}$$

and $\boldsymbol{u}(\ell)$ are the noise reference signals at the output of the BM. Substituting the constraints set of (3.43) in (3.55) yields:

$$y(\ell) = \sum_{p=1}^{P} g_p^* h_{\text{ref}}^p s^p(\ell) + \boldsymbol{w}^H \boldsymbol{v}(\ell). \tag{3.57}$$

Note that the output of the GSC in the transformed domain and, by equivalence, the LCMV in the original domain, is comprised of a summation of filtered versions of the P sources and a residual noise component. It is interesting to compare the different combinations of the constrained sources at the output of the regular LCMV-BF (3.28) and the extended GSC-BF (3.57).

In conclusion, applying a transformation $\boldsymbol{U}^H$ that preserves the rank-$\bar{M}$ signal subspace, guarantees the equivalence between the LCMV-BFs in the original and the transformed domains. Furthermore, an equivalent extended GSC structure exists in the transformed domain. Its optimality can be guaranteed by designing a FBF (3.51a) which satisfies the transformed constraints set (3.43), and by designing a BM (3.51b) with $\bar{M} - P$ linearly independent noise references.

In the following section we propose a specific transformation $\boldsymbol{U}$ which enables the construction of a distributed version of the extended GSC-BF.

3.2.3 DGSC

A recursive distributed version of the GSC-BF is now proposed. We present a specific transformation matrix $\boldsymbol{U}$ which conveniently splits the centralized GSC into a sum of N GSC-BFs, denoted $\boldsymbol{w}_n$ for $n \in \{1, \ldots, N\}$, operating in each of the WASN nodes. The proposed transformation matrix consists of N sub-matrices:

$$\boldsymbol{U} \triangleq \begin{bmatrix} \boldsymbol{U}_1 & \cdots & \boldsymbol{U}_N \end{bmatrix} \tag{3.58}$$

where the *transformed* inputs of the nth node are constructed by

$$\boldsymbol{z}_n(\ell) \triangleq \boldsymbol{U}_n^H \bar{\boldsymbol{z}}(\ell) \tag{3.59}$$

and the concatenation of all transformed inputs yields:

$$\boldsymbol{z}(\ell) \triangleq \left[\begin{array}{ccc} \boldsymbol{z}_1^T(\ell) & \cdots & \boldsymbol{z}_N^T(\ell) \end{array} \right]^T. \tag{3.60}$$

Note that $\boldsymbol{U}_n$ is an $\bar{M} \times M_n$ matrix and the corresponding transformed input $\boldsymbol{z}_n(\ell)$ is an $M_n \times 1$ vector for $n = 1, \ldots, N$. The sub-matrices $\boldsymbol{U}_n;\ n = 1, \ldots, N$ will be later defined.

The transformed inputs of each node will comprise all of its local microphone signals and a subset of the P shared signals. With the proposed transformation each node has at least P input signals, allowing for P constraints to be maintained locally, without unnecessary sacrificing degrees of freedom, as will be shown in the following sub-sections. In this section, the selection of the P shared signals is arbitrary, and should only satisfy linear independence. We will elaborate on this matter in Sec. 3.2.3.3. A specific and simple selection of the P shared signals is given in Sec. 3.2.4. The N outputs of the GSC-BFs, denoted $y_n(\ell)$ for $n = 1, \ldots, N$, and the P shared signals are transmitted in the WASN, where:

$$y_n(\ell) \triangleq \boldsymbol{w}_n^H \boldsymbol{z}_n(\ell) \tag{3.61}$$

and $\boldsymbol{w}_n$ is the GSC-BF at the nth node. Hence, a total of $N + P$ transmission channels are required by the algorithm. These channels effectively extend the number of available microphones at each node and should be continuously broadcast (also after the algorithm has converged). Note that for a node n that comprises a single microphone, i.e., $\bar{M}_n = 1$, no communication-bandwidth reduction is obtained, since the single microphone signal is transmitted. The global GSC-BF is given by augmenting the N nodes' BFs:

$$\boldsymbol{w} \triangleq \left[\begin{array}{ccc} \boldsymbol{w}_1^T & \cdots & \boldsymbol{w}_N^T \end{array} \right]^T. \tag{3.62}$$

The final output of the algorithm is obtained by substituting (3.60), (3.61) and (3.62) in (3.55):

$$\begin{aligned} y(\ell) &= \sum_{n=1}^{N} \boldsymbol{w}_n^H \boldsymbol{z}_n(\ell) \\ &= \sum_{n=1}^{N} y_n(\ell). \end{aligned} \tag{3.63}$$

The GSC-BF at the nth node is given by:

$$\boldsymbol{w}_n \triangleq \boldsymbol{q}_n - \boldsymbol{B}_n \boldsymbol{f}_n \tag{3.64}$$

where $\boldsymbol{q}_n$, $\boldsymbol{B}_n$ and $\boldsymbol{f}_n$ are the FBF, BM and NC at each node. Substituting (3.64) in (3.63), the output of the algorithm can be restated as:

$$y(\ell) = \sum_{n=1}^{N} (\boldsymbol{q}_n - \boldsymbol{B}_n \boldsymbol{f}_n)^H \boldsymbol{z}_n(\ell). \tag{3.65}$$

Considering (3.65), we identify the global components of the GSC-BF (3.50) as a concatenation of $\boldsymbol{q}_n$ and $\boldsymbol{f}_n$ for $n \in \{1, \ldots, N\}$, respectively:

$$\boldsymbol{q} \triangleq \begin{bmatrix} \boldsymbol{q}_1^T & \cdots & \boldsymbol{q}_N^T \end{bmatrix}^T \tag{3.66a}$$

$$\boldsymbol{f} \triangleq \begin{bmatrix} \boldsymbol{f}_1^T & \cdots & \boldsymbol{f}_N^T \end{bmatrix}^T. \tag{3.66b}$$

The global BM, $\boldsymbol{B}$, is constructed as a block-diagonal matrix with N blocks:

$$\boldsymbol{B} \triangleq \text{blkdiag}\left\{ \boldsymbol{B}_1, \quad \cdots, \quad \boldsymbol{B}_N \right\}. \tag{3.67}$$

Similarly to the notation in (3.55), (3.56a), (3.56b), (3.56c) for the global GSC, the outputs of the upper and lower branches, and the noise references at the nth node, are defined as:

$$y_n(\ell) \triangleq y_n^{\text{FBF}}(\ell) - y_n^{\text{NC}}(\ell) \tag{3.68a}$$

$$y_n^{\text{FBF}}(\ell) \triangleq \boldsymbol{q}_n^H \boldsymbol{z}_n(\ell) \tag{3.68b}$$

$$y_n^{\text{NC}}(\ell) \triangleq \boldsymbol{f}_n^H \boldsymbol{u}_n(\ell) \tag{3.68c}$$

$$\boldsymbol{u}_n(\ell) \triangleq \boldsymbol{B}_n^H \boldsymbol{z}_n(\ell). \tag{3.68d}$$

The global noise references vector is given by augmenting the noise reference signals of all nodes:

$$\boldsymbol{u}(\ell) \triangleq \begin{bmatrix} \boldsymbol{u}_1^T(\ell) & \cdots & \boldsymbol{u}_N^T(\ell) \end{bmatrix}^T. \tag{3.69}$$

A proper selection of P shared signals ensures that the number of noise references at the

output of the global BM is $\bar{M}-P$, and hence satisfies the requirement that $\boldsymbol{B}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}$ is a full-rank $(\bar{M}-P)\times(\bar{M}-P)$ matrix.

In the following, we prove analytically that the proposed DGSC converges to the centralized GSC. In Sec. 3.2.3.1 we propose a proper transformation matrix $\boldsymbol{U}$, that will allow us to split the BF into the structure defined by (3.65). We show that the proposed transformation matrix preserves the rank-$\bar{M}$ signals subspace, as required for the equivalence shown in Sec. 3.2.2. The design of FBF, BM, and NC of the DGSC is presented in Secs. 3.2.3.2,3.2.3.3,3.2.3.4. This structure is shown to satisfy the requirements of Sec. 3.2.2.

3.2.3.1 The transformation matrix

In the following, we define some notations for formulating the DGSC. The node that transmits the shared signal of the pth speaker is denoted as the "owner" of the pth source. In Sec. 3.2.4 we describe the procedure for selecting the owners of each of the P signals[2], and for generating the shared signals. Denote by $\chi(p)$ the index of the node which is the owner of the pth source. The shared signals are denoted by $r^p(\ell);\ p=1,\ldots,P$. Consider the pth shared signal, corresponding to the pth source. Assume that the pth source is owned by the nth node, i.e. $\chi(p)=n$. We suggest to construct the shared signal as:

$$\begin{aligned} r^p(\ell) &\triangleq \left(\boldsymbol{d}_n^p\right)^H \bar{\boldsymbol{z}}_n(\ell) \\ &= \left(\boldsymbol{d}_n^p\right)^H \boldsymbol{T}_n^H \left(\sum_{p=1}^{P} \bar{\boldsymbol{h}}^p s^p\left(\ell\right) + \bar{\boldsymbol{v}}\left(\ell\right)\right) \end{aligned} \tag{3.70}$$

where $\boldsymbol{d}_n^p$ is an $\bar{M}_n\times 1$ "local" BF that processes only the microphone signals of the nth node. A specific choice of the BFs $\boldsymbol{d}_n^p$ for $p=1,\ldots,P$ and $n=\chi(p)$ will be defined in Sec. 3.2.4.

Denote by $\mathcal{P}_n \triangleq \left\{\ p_n(1),\ \cdots,\ p_n(P_n)\ \right\}$ the set of sources owned by the nth node, and by $P_n \triangleq |\mathcal{P}_n|$ the number of sources owned by the nth node. The shared signals generated by the nth node, are defined in a vector notation by the $P_n\times 1$ vector:

$$\boldsymbol{r}_n(\ell) \triangleq \left[\ r^{p_n(1)}(\ell)\ \ \cdots\ \ r^{p_n(P_n)}(\ell)\ \right]^T \tag{3.71}$$

$$= \boldsymbol{D}_n^H \bar{\boldsymbol{z}}_n(\ell) \tag{3.72}$$

[2] A node can be the owner of several sources.

where

$$\boldsymbol{D}_n \triangleq \left[\begin{array}{ccc} \boldsymbol{d}_n^{p_n(1)} & \cdots & \boldsymbol{d}_n^{p_n(P_n)} \end{array} \right]. \tag{3.73}$$

The $\bar{M}_n \times P_n$ dimensional matrix $\boldsymbol{D}_n$ should be properly constructed to have a rank P_n. As each source is exclusively owned by a single node

$$P = \sum_{n=1}^{N} P_n. \tag{3.74}$$

The $P \times 1$ vector of all shared signals is constructed by augmenting the contributions of all nodes:

$$\boldsymbol{r}(\ell) \triangleq \left[\begin{array}{ccc} \boldsymbol{r}_1^T(\ell) & \cdots & \boldsymbol{r}_N^T(\ell) \end{array} \right]^T. \tag{3.75}$$

Note, that some of the nodes may own no sources. For instance, suppose that the n'th node does not own any source. In that case, $P_{n'} = 0$ and the corresponding vector of shared signals $\boldsymbol{r}_{n'}(\ell)$ will be empty.

The set of indexes of the P speakers is denoted by $\mathcal{P} \triangleq \{1, \ldots, P\}$. Denote the set of shared signals that the nth node receives as $\dot{\mathcal{P}}_n$. It comprises the indexes of all sources except the self-owned sources:

$$\begin{aligned} \dot{\mathcal{P}}_n \triangleq & \mathcal{P} \backslash \mathcal{P}_n \\ = & \left\{ \begin{array}{ccc} \dot{p}_n(1) & \cdots & \dot{p}_n(\dot{P}_n) \end{array} \right\} \end{aligned} \tag{3.76}$$

where $\backslash$ denotes the set subtraction operation and $\dot{P}_n = |\dot{\mathcal{P}}_n|$. The $\dot{P}_n \times 1$ vector of shared signals received by the nth node is denoted by:

$$\begin{aligned} & \dot{\boldsymbol{r}}_n^T(\ell) \triangleq \\ & \quad \left[\begin{array}{ccccccc} \boldsymbol{r}_1^T(\ell) & \cdots & \boldsymbol{r}_{n-1}^T(\ell) & \boldsymbol{r}_{n+1}^T(\ell) & \cdots & \boldsymbol{r}_N^T(\ell) \end{array} \right]. \end{aligned} \tag{3.77}$$

As previously defined in (3.59), the signals available for processing at the nth node are denoted by $\boldsymbol{z}_n(\ell)$, an $M_n \times 1$ vector:

$$\boldsymbol{z}_n(\ell) \triangleq \boldsymbol{U}_n^H \bar{\boldsymbol{z}}(\ell)$$

where

$$\boldsymbol{U}_n \triangleq \begin{bmatrix} \boldsymbol{T}_n & \dot{\boldsymbol{T}}_n \end{bmatrix} \tag{3.78a}$$

$$\dot{\boldsymbol{T}}_n \triangleq \begin{bmatrix} \boldsymbol{T}_1\boldsymbol{D}_1 & \cdots & \boldsymbol{T}_{n-1}\boldsymbol{D}_{n-1} \\ \boldsymbol{T}_{n+1}\boldsymbol{D}_{n+1} & \cdots & \boldsymbol{T}_N\boldsymbol{D}_N \end{bmatrix}. \tag{3.78b}$$

From (3.78a), the number of transformed input signals at the nth node is given by:

$$M_n = \bar{M}_n + \dot{P}_n. \tag{3.79}$$

Note that $\boldsymbol{T}_n$ and $\boldsymbol{T}_{n'}$ $\forall n \neq n'$ are linearly independent, since they comprise different microphones. Now, since the rank of $\boldsymbol{D}_n$ in (3.73) is P_n, it follows that the rank of $\boldsymbol{T}_n\boldsymbol{D}_n$ is also P_n. Hence, we argue that the rank of $\dot{\boldsymbol{T}}_n$ is $\sum_{n'\neq n} P_{n'} = \dot{P}_n$. A similar argument can be applied to $\boldsymbol{U}_n$. Constructed as a concatenation of $\boldsymbol{T}_n$ and $\dot{\boldsymbol{T}}_n$, its rank equals M_n.

We designate the pth shared signal, $r^p(\ell)$, as the reference microphone for the pth source RTF (3.41). We identify the ATF relating the pth source and the pth shared signal (3.70) as:

$$h^p_{\text{ref}} \triangleq \left(\boldsymbol{d}^p_n\right)^H \boldsymbol{T}_n^H \bar{\boldsymbol{h}}^p. \tag{3.80}$$

Now, the pth RTF (3.41) can be defined with respect to the pth shared signal. Considerations for constructing $\boldsymbol{d}^p_n;\ p = 1, \ldots, P$ will be discussed in Sec. 3.2.4.

The proposed $\bar{M} \times M$ dimensional transformation matrix is finally given by:

$$\boldsymbol{U} \triangleq \begin{bmatrix} \boldsymbol{U}_1 & \cdots & \boldsymbol{U}_N \end{bmatrix} \tag{3.81}$$

where we note that

$$\begin{aligned} M =& \sum_{n=1}^{N} M_n = \sum_{n=1}^{N} \left(\bar{M}_n + \dot{P}_n\right) \\ =& \bar{M} + (N-1)\, P. \end{aligned} \tag{3.82}$$

It can be easily shown that the rank of the column-subspace of $\boldsymbol{U}$ is $\bar{M}$, since a permutation of $\begin{bmatrix} \boldsymbol{T}_1 & \cdots & \boldsymbol{T}_N \end{bmatrix} = \boldsymbol{I}_{\bar{M}}$ is its sub-matrix. Hence, $\boldsymbol{U}$ is a valid transformation matrix, rendering $\boldsymbol{w}$ and $\bar{\boldsymbol{w}}_{\text{LCMV}}$ equivalent (3.40).

According to (3.59) and (3.78a), the transformed input vector in the nth node is the

M_n-dimensional vector:

$$\boldsymbol{z}_n(\ell) = \left[\begin{array}{cc} \bar{\boldsymbol{z}}_n^T(\ell) & \dot{\boldsymbol{r}}_n^T(\ell) \end{array} \right]^T \tag{3.83}$$

where the received shared signals at the nth node are given by the $\dot{P}_n \times 1$ vector $\dot{\boldsymbol{r}}_n(\ell) \triangleq \dot{\boldsymbol{T}}_n^H \bar{\boldsymbol{z}}(\ell)$.

Examining (3.19) and (3.59), the transformed inputs vector of the nth node is given by:

$$\boldsymbol{z}_n(\ell) = \tilde{\boldsymbol{H}}_n \boldsymbol{s}(\ell) + \boldsymbol{v}_n(\ell) \tag{3.84}$$

where $\tilde{\boldsymbol{H}}_n = \boldsymbol{U}_n^H \bar{\boldsymbol{H}}$ is an $M_n \times P$ matrix and $\boldsymbol{v}_n(\ell) = \boldsymbol{U}_n^H \bar{\boldsymbol{v}}(\ell)$.

Define

$$\boldsymbol{H}_r \triangleq \boldsymbol{D}^H \bar{\boldsymbol{H}} \tag{3.85}$$

where $\boldsymbol{D}$ is defined as

$$\boldsymbol{D} \triangleq \left[\begin{array}{ccc} \boldsymbol{T}_1 \boldsymbol{D}_1 & \cdots & \boldsymbol{T}_N \boldsymbol{D}_N \end{array} \right] \tag{3.86}$$

and $\boldsymbol{D}_n$ is defined in (3.73). The elements of the $P \times P$ matrix $\boldsymbol{H}_r$ are the ATFs relating the speakers and the P shared signals. We assume that $\boldsymbol{H}_r$ is a full rank matrix. The condition for the invertibility of $\boldsymbol{H}_r$ is given is Sec. 3.2.4.

We now show that the rank of $\tilde{\boldsymbol{H}}_n$ is P for $n = 1, \ldots, N$. Notice that the matrix $\boldsymbol{H}_r$ is a column permutation of the $P \times P$ matrix:

$$\left[\begin{array}{c|c} \boldsymbol{D}_n & \boldsymbol{0}_{\bar{M}_n \times \dot{P}_n} \\ \hline \boldsymbol{0}_{\dot{P}_n \times P_n} & \boldsymbol{I}_{\dot{P}_n \times \dot{P}_n} \end{array} \right]^H \tilde{\boldsymbol{H}}_n.$$

Since $P = \operatorname{rank}\{\boldsymbol{H}_r\} \leq \operatorname{rank}\left\{\tilde{\boldsymbol{H}}_n\right\} \leq P$, we conclude that $\operatorname{rank}\left\{\tilde{\boldsymbol{H}}_n\right\} = P$.

Determining $\boldsymbol{U}$ as above is instrumental for transforming, the centralized GSC-BF into a sum of N GSC-BFs in the transformed domain. The total output of the DGSC algorithm is available at each of the nodes in the WASN.

In the following sections we substitute $\tilde{\boldsymbol{H}}_n$ by the RTFs matrix

$$\boldsymbol{H}_n \triangleq \boldsymbol{U}_n^H \tilde{\boldsymbol{H}} \boldsymbol{H}_r^{-1} \tag{3.87}$$

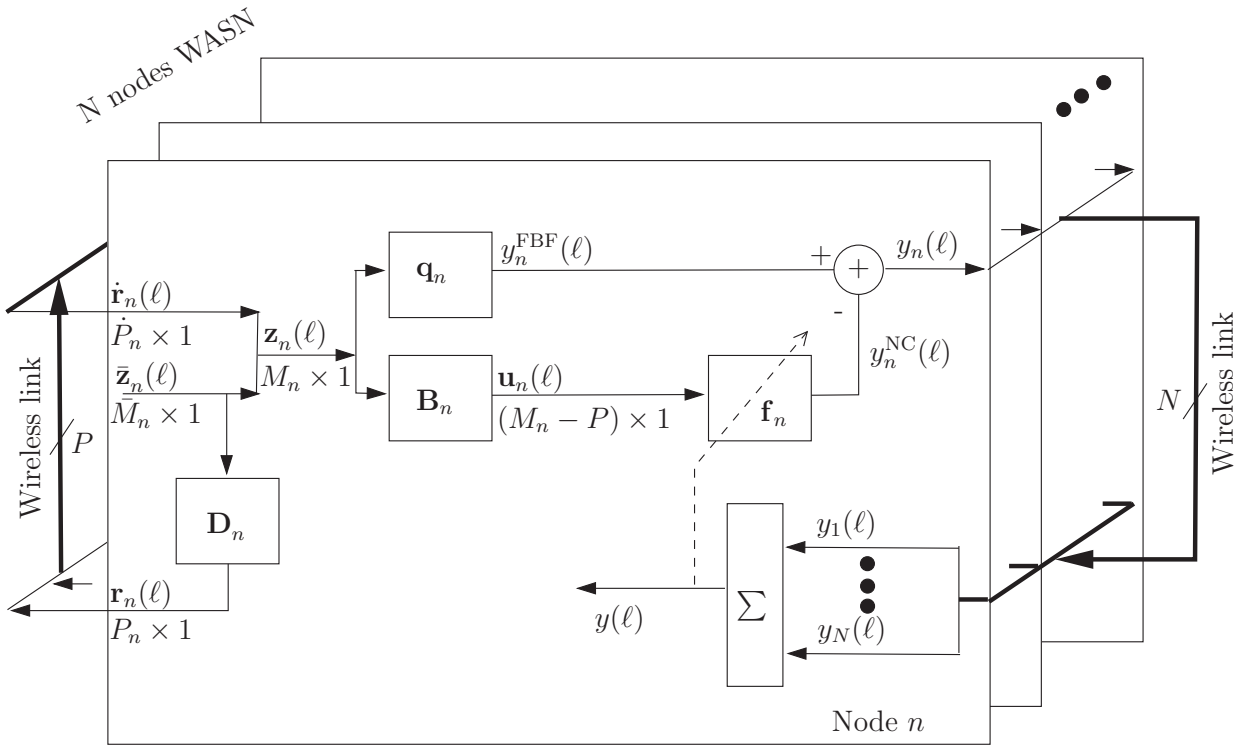


Figure 3.4: The DGSC.

for $n = 1, \ldots, N$. A block-diagram of the proposed algorithm is depicted in Fig. 3.4.

3.2.3.2 The distributed FBF

Had $\boldsymbol{H}$ been known to all nodes, it would have been possible to calculate the classic centralized FBF, $\boldsymbol{q}_{\mathrm{GSC}}$. In our case, we propose a distributed FBF consisting of a summation of local FBFs, which are calculated from the transformed RTFs at each node. Explicitly, the proposed distributed FBF at the nth node is defined as:

$$\boldsymbol{q}_n \triangleq \frac{1}{N} \boldsymbol{H}_n \left(\boldsymbol{H}_n^H \boldsymbol{H}_n \right)^{-1} \boldsymbol{g}. \tag{3.88}$$

As $\boldsymbol{H}_n$ equals $\tilde{\boldsymbol{H}}_n$ up to a different column scaling, its rank equals P. Therefore, $\boldsymbol{H}_n^H \boldsymbol{H}_n$ is an invertible matrix. As stated earlier, the FBF is not unique, and can have different forms with different selections of $\boldsymbol{a}, \boldsymbol{q}^{uc}, \boldsymbol{B}^{uc}$ in (3.51a),(3.51b). Various choices of the FBF will differ in their robustness to estimation errors.

It can be easily verified, by substituting (3.88) in (3.66a), that the global distributed

FBF (3.51a) satisfies the global constraints set (3.43) since

$$\begin{aligned}\boldsymbol{H}^H\boldsymbol{q} =& \left(\sum_{n=1}^{N}\boldsymbol{H}_n^H\boldsymbol{q}_n\right)\\ =&\frac{1}{N}\sum_{n=1}^{N}\boldsymbol{g}\\ =&\boldsymbol{g}.\end{aligned} \tag{3.89}$$

This simple FBF design utilizes each of the WASN microphones and is not optimal in any sense. The robustness analysis of the proposed algorithm to estimation errors is out of the scope of the current contribution.

3.2.3.3 The distributed BM

As mentioned earlier, the BM is not unique, and several procedures for its construction are available. Recently, we have proposed an efficient implementation of a sparse BM [101]. Similarly to the construction of the BM in Sec. 3.2.2, we propose that the nth node will construct a transformed BM by applying the SVD to $\boldsymbol{H}_n$, for $n = 1, \ldots, N$. The SVD of $\boldsymbol{H}_n$ is

$$\boldsymbol{H}_n = \begin{bmatrix} \boldsymbol{\Gamma}_n & \boldsymbol{B}_n \end{bmatrix} \begin{bmatrix} \boldsymbol{\Lambda}_n \\ \boldsymbol{0}_{(M_n-P)\times P} \end{bmatrix} \boldsymbol{\Theta}_n^H. \tag{3.90}$$

where the column-space of $\boldsymbol{H}_n$ is spanned by the column-space of $\boldsymbol{\Gamma}_n$. The null-subspace of $\boldsymbol{H}_n$ is spanned by the column-subspace of $\boldsymbol{B}_n$ and hence is an adequate BM at the nth node. Since the column rank of $\boldsymbol{H}_n$ is P, The dimensions of the BM at the nth node are $M_n \times (M_n - P)$, and its column rank is $M_n - P$.

Next, we prove that $\boldsymbol{B}$ is a valid BM. From its construction (3.67), it trivially blocks $\boldsymbol{H}$, hence, in order to complete the proof, we need to show that $\boldsymbol{B}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}$ is of full-rank. From the definition of $\boldsymbol{\Phi}_{vv}$ in (3.30b), and since $\bar{\boldsymbol{\Phi}}_{vv}$ is full-rank (rank-$\bar{M}$), the latter condition is equivalent to showing that the column rank of $\boldsymbol{U}\boldsymbol{B} = \begin{bmatrix} \boldsymbol{U}_1\boldsymbol{B}_1 & \cdots & \boldsymbol{U}_N\boldsymbol{B}_N \end{bmatrix}$ is $\bar{M} - P$.

The rank of $\boldsymbol{U}_n$ is M_n. A one-to-one linear transformation from $\boldsymbol{U}_n$ to $\begin{bmatrix} \boldsymbol{Q}_n & \bar{\boldsymbol{H}} \end{bmatrix}$ exists for $n = 1, \ldots, N$ where $\boldsymbol{Q}_n$ is an $\bar{M} \times (M_n - P)$ matrix orthogonal to $\bar{\boldsymbol{H}}$. It follows that $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_N$ share at least P degrees of freedom (the columns of $\bar{\boldsymbol{H}}$). Now, since the rank of $\boldsymbol{U}$ is $\bar{M}$, we conclude that $\boldsymbol{U}_1, \ldots, \boldsymbol{U}_N$ share exactly P degrees of freedom. Hence, the

rank of $\begin{bmatrix} \boldsymbol{Q}_1 & \cdots & \boldsymbol{Q}_N \end{bmatrix}$ is $\bar{M} - P$. By construction, $\boldsymbol{Q}_n$ is an $\bar{M} \times (M_n - P)$ BM of $\bar{\boldsymbol{H}}$, and its $M_n - P$ outputs $\boldsymbol{Q}_n^H \bar{\boldsymbol{z}}(\ell)$ are equivalent (represent the same noise signals) to the $M_n - P$ outputs $\boldsymbol{B}_n^H \boldsymbol{z}_n(\ell) = \boldsymbol{B}_n^H \boldsymbol{U}_n^H \bar{\boldsymbol{z}}(\ell)$. Finally, $\boldsymbol{U}\boldsymbol{B}$ is a concatenation of the sub-matrices $\boldsymbol{U}_n \boldsymbol{B}_n$ for $n = 1, \ldots, N$. Hence, it has the same rank as the concatenation of $\boldsymbol{Q}_n$ for $n = 1, \ldots, N$. Based on the above discussion, it is guaranteed that $\boldsymbol{B}^H \boldsymbol{\Phi}_{vv} \boldsymbol{B}$ is a full-rank $\left(\bar{M} - P\right) \times \left(\bar{M} - P\right)$ matrix.

3.2.3.4 The distributed NC

The normalized LMS (NLMS) adaptation of the global NC in [39] is given by:

$$\boldsymbol{f}(\ell) = \boldsymbol{f}(\ell - 1) + \mu \frac{\boldsymbol{u}(\ell) y^*(\ell)}{P_u(\ell)} \tag{3.91}$$

where $P_u(\ell)$ is a recursive estimator of the power of the noise reference signals, i.e., $\mathrm{E}\left\{\|\boldsymbol{u}(\ell)\|^2\right\}$:

$$P_u(\ell) = \rho P_u(\ell - 1) + (1 - \rho)\|\boldsymbol{u}(\ell)\|^2 \tag{3.92}$$

where ρ is a forgetting factor (typically $0.8 < \rho < 1$). Due to inevitable estimation errors, some of the speech signals might leak to the noise reference signals. In order to prevent the self-cancelation phenomenon, which is manifested in a severe speech distortion, the NC is updated according to (3.91) only when the speakers are inactive. A perfect VAD is assumed for this purpose. The total output of the algorithm, $y(\ell)$, is available to all nodes as the summation in (3.65). As clearly seen in (3.91), the noise reference signals at the nth node, $\boldsymbol{u}_n(\ell)$, only affect $\boldsymbol{f}_n(\ell)$. Hence, updating the NC is equivalent to N simultaneous updates of the distributed NCs $\boldsymbol{f}_n(\ell)$; $n = 1, \ldots, N$. Explicitly, the recursive update of the distributed NC is given by:

$$\boldsymbol{f}_n(\ell) = \boldsymbol{f}_n(\ell - 1) + \mu \frac{\boldsymbol{u}_n(\ell) y^*(\ell)}{P_{u,n}(\ell)} \tag{3.93}$$

where $P_{u,n}(\ell)$ is the estimated power of the global noise reference vector $\mathrm{E}\left\{\|\boldsymbol{u}(\ell)\|^2\right\}$ at the nth node. We assume that the power of the local noise reference signals at the various nodes are approximately the same, i.e $\mathrm{E}\left\{\|\boldsymbol{u}(\ell)\|^2\right\} = \frac{\bar{M}-P}{M_n-P_n} \mathrm{E}\left\{\|\boldsymbol{u}_n(\ell)\|^2\right\}$; $n = 1, \ldots, N$. Hence

the estimated power at the nth node is:

$$P_{u,n}(\ell) = \rho P_{u,n}(\ell - 1) + (1 - \rho)\frac{\bar{M} - P}{M_n - P_n}\|\boldsymbol{u}_n(\ell)\|^2. \tag{3.94}$$

The latter assumption can be circumvented by sharing estimates of the variance of the noise reference signals $\boldsymbol{u}_n(\ell);\ n = 1, \ldots, N$ in the WASN. Assuming that the noise statistics is slowly varying, the latter exchange of power estimates does not consume a large bandwidth.

3.2.4 Shared signals construction

Here, we propose a simple procedure for generating the shared signals, which is based on selecting the microphones with the highest SNR for each of the sources. Since the pth shared signal is used as the reference signal in the definition of the RTF (3.80), and since in practice the RTF is unknown and has to be estimated, it is desired that the SNR of the pth source will be maximal. The SNR of a microphone with respect to some source p is defined as the ratio between the source power and the power of the slowly time-varying noise.

As mentioned in Sec. 3.2.3, the shared signals should satisfy that the column rank of $\boldsymbol{H}_r$ is P. Therefore, a microphone that was selected as the shared signal of a certain source, cannot be chosen as a shared signal for another source, or else the rank of $\boldsymbol{H}_r$ will be lower than P.

During the initialization of the algorithm each node sets $\mathcal{J}_n \triangleq \{1, \ldots, \bar{M}_n\}$ the index set of candidate microphones for shared signals. For each source $p \in \{1, 2, \ldots, P\}$ the following procedure is applied. First, the nth node estimates $\gamma_n^p(j);\ j \in \mathcal{J}_n$, the pth source SNR at each of its available local microphones, $m_n(j);\ j \in \mathcal{J}_n$. Each node selects the microphone with the highest SNR. The SNR and the index of the candidate microphone of the nth node are:

$$\gamma_n^p \triangleq \max_{j \in \mathcal{J}_n} \gamma_n^p(j) \tag{3.95a}$$

$$j_n^p \triangleq \operatorname*{argmax}_{j \in \mathcal{J}_n} \gamma_n^p(j). \tag{3.95b}$$

Each node shares the maximal SNR γ_n^p with the rest of the nodes.

The node n' with the maximum SNR will be declared the owner of the source p, i.e.,

$\chi(p) = n'$:

$$\chi(p) \triangleq \underset{n=1,\ldots,N}{\operatorname{argmax}} \gamma_n^p. \tag{3.96}$$

The n'th node constructs the BF that extracts the pth shared signal

$$\boldsymbol{d}_{n'}^p \triangleq \left[\begin{array}{ccc} \mathbf{0}_{1\times(j_{n'}^p-1)} & 1 & \mathbf{0}_{1\times(\bar{M}_{n'}-j_{n'}^p)} \end{array} \right]^T \tag{3.97}$$

and removes $j_{n'}^p$ from its set of candidate microphones to own a signal

$$\mathcal{J}_{n'} = \mathcal{J}_{n'} \backslash j_{n'}^p. \tag{3.98}$$

This way, it is guaranteed that a single microphone will not be chosen more than once. The procedure is repeated for all sources, resulting in the entire set of shared signals. Note that some nodes may be the owners of more than a single source, and some nodes may have no ownership on sources. The proposed method is very simple, and does not require any processing for constructing the shared signals. In practice, $\boldsymbol{H}_r$ is usually full-rank, however, this is not guaranteed. In case, that $\boldsymbol{H}_r$ is rank-deficient, a simple procedure of replacing some of the shared signals until the rank is full can be applied.

3.2.5 A comparison between the DGSC and the LC-DANSE

We compare the proposed DGSC and the LC-DANSE [74]. Both algorithms converge to the centralized LCMV-BF. The LC-DANSE implements a distributed version of the closed-form LCMV, whereas the DGSC adopts the GSC implementation of the LCMV structure. In the DGSC a common objective to all nodes, i.e. the classification of desired and competing speakers, yields a single common constraints set. A more general approach is adopted by the LC-DANSE, which allows node-specific constraint sets. In practice, this enables each node to define its own objective, i.e. a set of desired and competing speakers. The LC-DANSE is an iterative algorithm (although, the iterations can be carried out recursively over time), while the DGSC is a time-recursive algorithm. The GSC structure conveniently decouples the task of noise reduction from the task of satisfying the constraints set. Hence, allowing the ANC to adjust to variations in the noise statistics. The DGSC requires $N+P$ transmission channels, whereas the LC-DANSE requires $N \times P$ transmission channels. Both algorithms, require estimates of the sources RTFs. In static scenarios, the DGSC requires a single estimate

thereof, whereas in the LC-DANSE, each iteration requires additional RTF estimates. In the following section, we experimentally compare the DGSC and the LC-DANSE.

3.2.6 Experimental study

In order to verify the equivalence between the centralized GSC and the proposed DGSC, a comprehensive experimental study is carried out. The validity of the proposed algorithm is tested for narrowband signals in Sec. 3.2.6.1 and for speech signals in Sec. 3.2.6.2. We compare the following five algorithms, namely, the centralized closed-form LCMV, the centralized GSC, a single node local GSC (arbitrarily chosen as the first node), the LC-DANSE and the proposed DGSC algorithm. The comparison criteria are noise reduction and distortion of the constrained sources. Opposed to the global BFs and the DGSC algorithms where the number of constraints can be as large as the total number of microphones in the WASN ($P \leq \bar{M}$), the local GSC is constrained to handle only scenarios where $P \leq \bar{M}_1$. The performance is averaged over multiple Monte-Carlo experiments in various scenarios.

3.2.6.1 Narrowband signals

A WASN comprising $N = 4$ nodes, each consisting of $\bar{M}_n = 4$ microphones was simulated. We denote by *constrained* sources, sources for which desired responses exist and are maintained with a proper linear constraints set. Furthermore, we denote by *unconstrained* sources, all interfering sources that $\bar{\boldsymbol{v}}(\ell, k)$ comprises. We examine a total of 28 scenarios: all combinations of $P = 1, 2, \ldots, 7$ constrained sources and $P_i = 1, 3, 5, 7$ unconstrained sources. A spatially white Gaussian sensor noise is added to the microphone signals. In each scenario (a specific selection of P and P_i), 10 sets of source ATFs and a vector of desired responses are randomized. For each set, 10 realizations of 10^5 samples of $P + P_i$ independent identically distributed (i.i.d.) Gaussian processes are randomized. These signals serve as the constrained and unconstrained sources. Note that in the narrowband case all sources are stationary. A total of 3200 Monte-Carlo experiments are used for the comparison of the various algorithms. The SNR, the ratio between the constrained signals power and the sensors spatially white noise, is set to 30dB, and the interference to noise ratio (INR), the ratio between the unconstrained sources power and the sensors noise, is set to 25dB. The step-size of the NLMS algorithms is set to $\mu = 0.25$. The results of the LC-DANSE algorithm are measured after 10 iterations. We assume that the RTFs are known without estimation errors, hence no distor-

tion to the constrained signals is measured for the centralized LCMV, the centralized GSC, and the DGSC for all values of $P \leq \bar{M} = 16$. For the single node GSC, there is no distortion for $P = 1, 3 \leq \bar{M}_1 = 4$, but for $P = 5, 7$, due to lack of degrees of freedom (there are only $\bar{M}_1 = 4$ beams that can be steered), distortion is inevitable. The distortion measured in the LC-DANSE is also low (< -23dB) in all scenarios. The NR of the various algorithms after convergence for $P_i = 3$ versus the number of constraints, P, is depicted in Fig. 3.6. The figure of merit is defined as the ratio between the slowly time varying noise power at the input and at the output. As expected, the NR of the centralized GSC is about 0.35dB lower than the centralized LCMV. This is a result of using the LMS algorithm, which suffers from excess MSE. It can be mitigated by reducing the step-size μ compromising convergence rate. The NR of the proposed DGSC is 0.52dB lower than the centralized GSC (probably since longer convergence time is required), whereas the of the NR single node GSC is much lower (from 7.7dB to 47.6dB, depending on the number of constraints). The NR performance of all BFs reduces as the number of constraints increases. The convergence of the NR versus the number of samples is depicted in Fig. 3.5 for a scenario with $P = 5$ and $P_i = 3$.

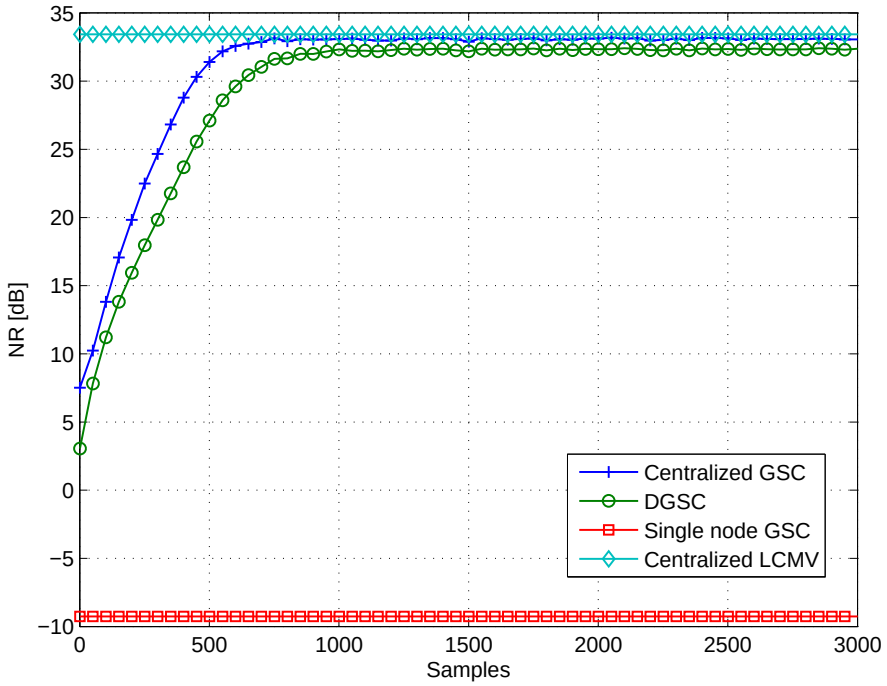


Figure 3.5: The convergence of the tested algorithms versus the number of samples for $P = 5$ constraints and $P_i = 3$.

Although the proposed DGSC and the centralized GSC converge to more or less the same NR as the centralized LCMV, the convergence time of the DGSC is higher. This may result from the higher condition number, defined as the ratio of the largest and smallest eigenvalues, of the noise references covariance matrix $\boldsymbol{B}^H\boldsymbol{\Phi}_{vv}\boldsymbol{B}$. Higher condition number is known to increase the convergence time [106]. For example, in the depicted scenario, the average condition number of the noise references covariance matrix of the DGSC is 6.9dB higher than of the centralized GSC. The latter phenomenon may be attributed to the vector $\boldsymbol{a} \neq \boldsymbol{0}$ in (3.51a), which increases the norm of the ANC in (3.51c), however, this subject requires further research.

The ratio of the noise level at the output of the DGSC and the noise level at the output of the centralized GSC is given in Table 3.1 for $P = 1, 2, \ldots, 8$ and $P_i = 1, 3, 5, 7$.

Table 3.1: The ratio of the noise level at the outputs of the DGSC and the centralized GSC [dB].

$P \backslash P_i$	1	3	5	7
1	0.03	0.20	0.18	0.26
2	0.08	0.22	0.37	0.44
3	0.32	0.22	0.39	0.62
4	0.25	0.56	0.48	0.75
5	0.51	0.50	0.89	0.66
6	0.38	0.90	0.45	1.01
7	0.15	1.29	0.91	0.95
8	0.31	0.75	0.90	0.84

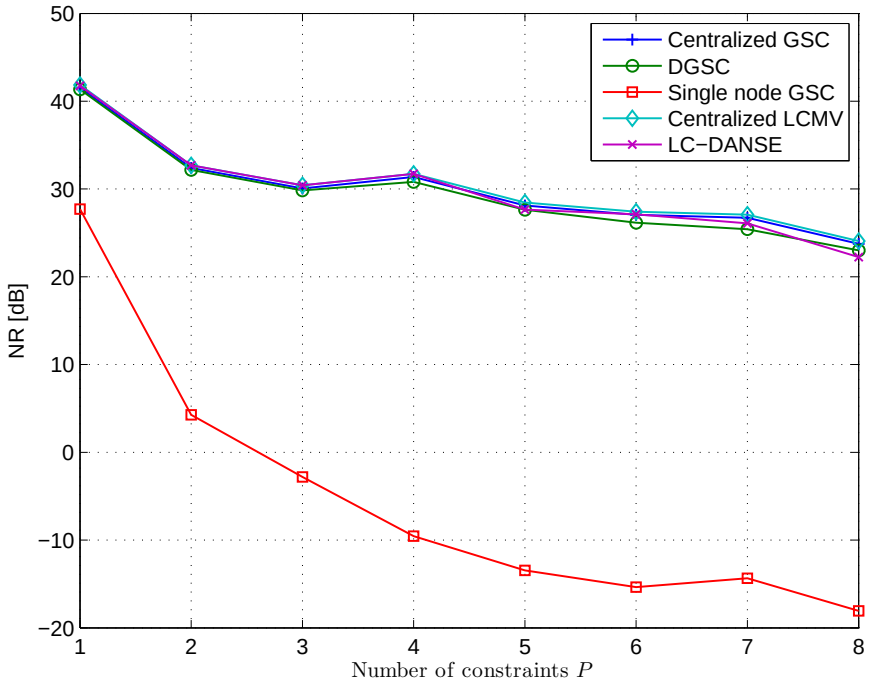


Figure 3.6: The NR of the tested algorithms versus the number of constraints P, for $P_i = 3$.

3.2.6.2 Speech signals

The performance of the various BFs is tested in a simulated room scenario, by using a RIR generator [107],[108]. The dimensions of the simulated room are 4m × 3m × 3m, and its reverberation time is set to $T_{60} = 300$ms. An $N = 4$ nodes WASN where each node comprises $\bar{M}_n = 2$ microphones at a distance of 5cm is set. The nodes are located at the center of each of the four walls, 10cm from the walls surface and at a height of 1.5m. A desired female speaker and a competing male speaker, are located in the room as well as two white Gaussian stationary interferences. The figures of merit of the BFs are tested by 90 Monte Carlo experiments, where in each experiment the sources locations are randomly selected, and the microphone constellation remains fixed. The room setup of one of the Monte Carlo experiments is depicted in Fig. 3.7.

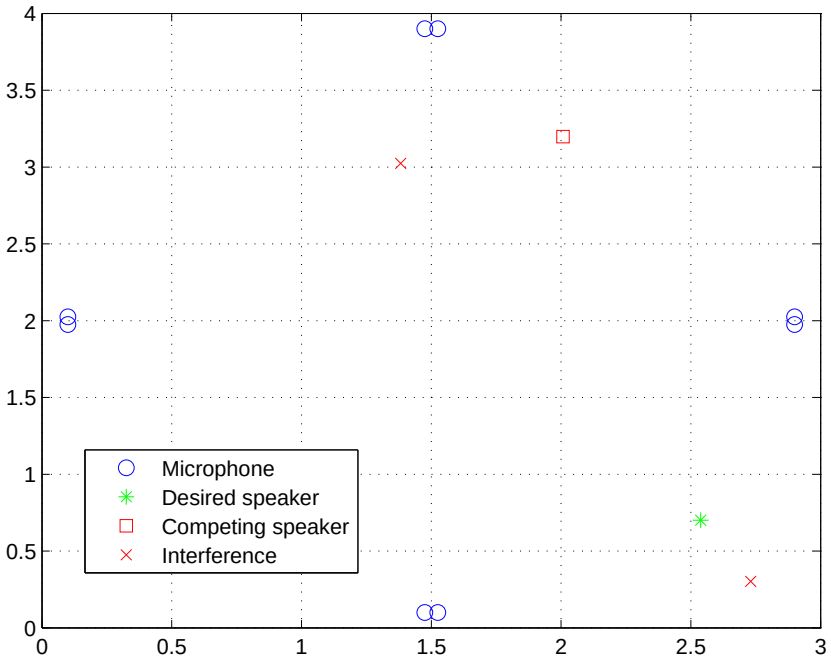


Figure 3.7: The room setup of one of the Monte Carlo simulations.

The microphone signals are sampled at a sampling rate of 8kHz. The length of the STFT window is 4096 points with 75% overlap between frames. The estimated RTFs are double sided filters 3072 coefficient long. They are estimated using the subspace method as in [51]. In the DGSC algorithm, in order to save communication-bandwidth, the signals undergo inverse STFT prior to the broadcast in the network. We use the overlap and save scheme for applying the filters in the STFT domain [109],[39]. The SNR improvement, SIR improvement and distortion measures of the centralized GSC, the DGSC and the single node GSC for the various Monte Carlo experiments are depicted in Figs. 3.8,3.9,3.10, respectively.

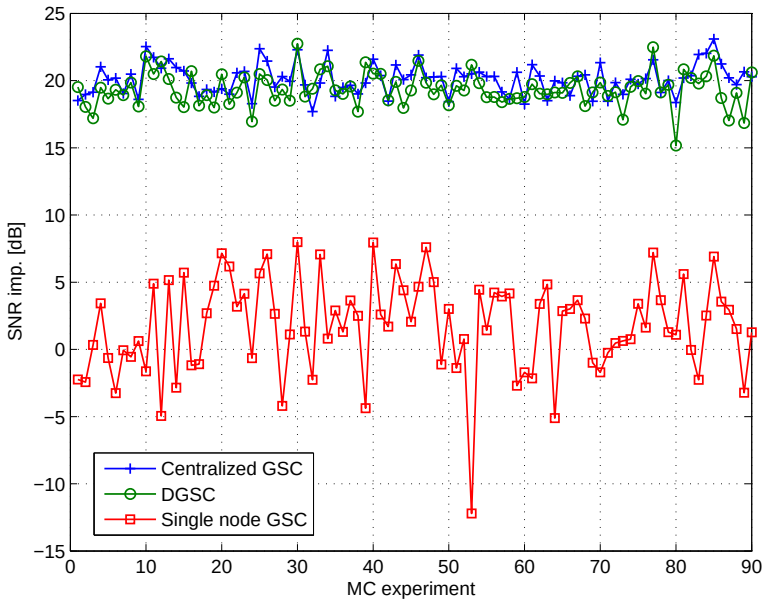


Figure 3.8: The SNR improvement of the tested algorithms in various Monte Carlo experiments.

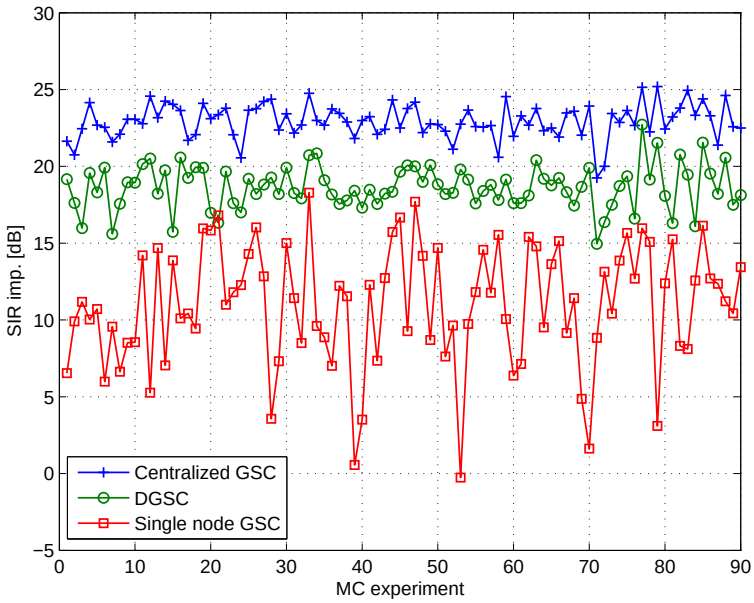


Figure 3.9: The SIR improvement of the tested algorithms in various Monte Carlo experiments.

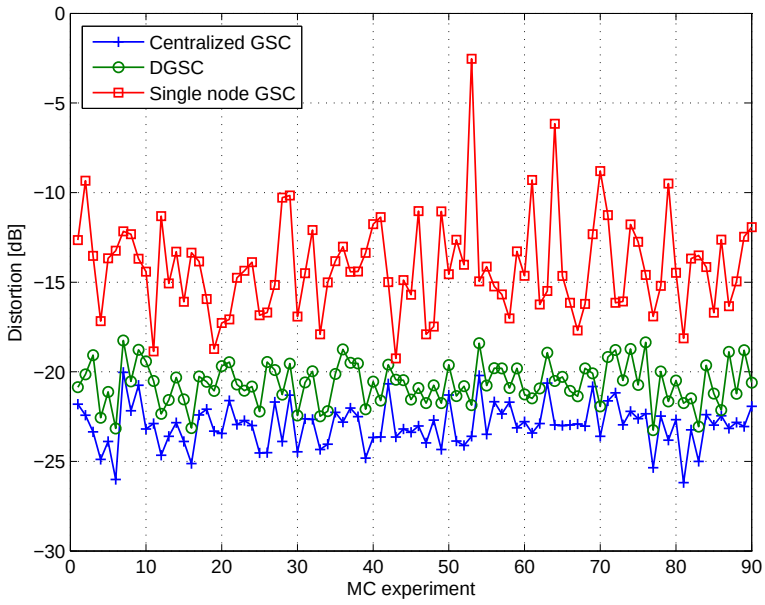


Figure 3.10: The distortion of the tested algorithms in various Monte Carlo experiments.

The SNR is the ratio between the powers of the desired speaker and the stationary noise, the SIR is the ratio between the powers of the desired speaker and the competing speaker, and the distortion is the ratio between the MSE of the desired speech at the output and the power of the desired speech signal. The SNR and the SIR at the input are set to 13dB and 0dB, respectively. It is clear from these figures that the NR values of the DGSC and the centralized GSC are equivalent, and that both outperform the single node GSC. The average figures of merit of the various algorithms are depicted in Table 3.2. The SNR improvement of the DGSC and the centralized GSC are similar (20.1dB and 19.3dB, respectively. The slight differences may be explained as in the narrowband case), while the SNR improvement of the single node GSC is significantly lower (1.7dB). The SIR improvement and the distortion of the centralized GSC are 22.9dB and -23.0dB, respectively, whereas the corresponding measures of the DGSC are a bit worse 18.6dB and -20.3dB, respectively. This may be attributed to differences in the robustness of the BFs against RTF estimation errors (see discussion in the narrowband case). Due to the significantly lower number of microphones, the SIR improvement and distortion of the single node GSC (11.0dB and -14.1dB, respectively) are much worse than the centralized GSC. The centralized GSC and the DGSC exhibit comparable convergence behaviour as depicted in Fig. 3.11.

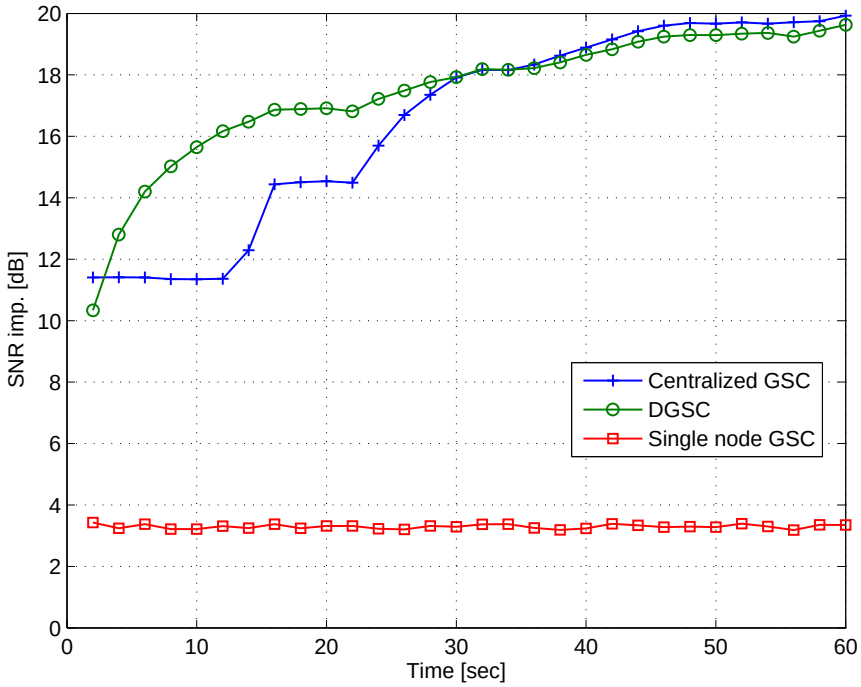


Figure 3.11: The convergence of the tested algorithms versus time.

Note, that the single node GSC converges much faster, but its overall performance is very poor.

Sonograms of the various components of the signal received in the first microphone, and the outputs of the centralized GSC, the DGSC and the single node GSC are depicted in Fig. 3.12. The equivalence of the DGSC and the centralized GSC and their superiority to the single node GSC can be deduced from the figures.

3.2.7 Conclusions

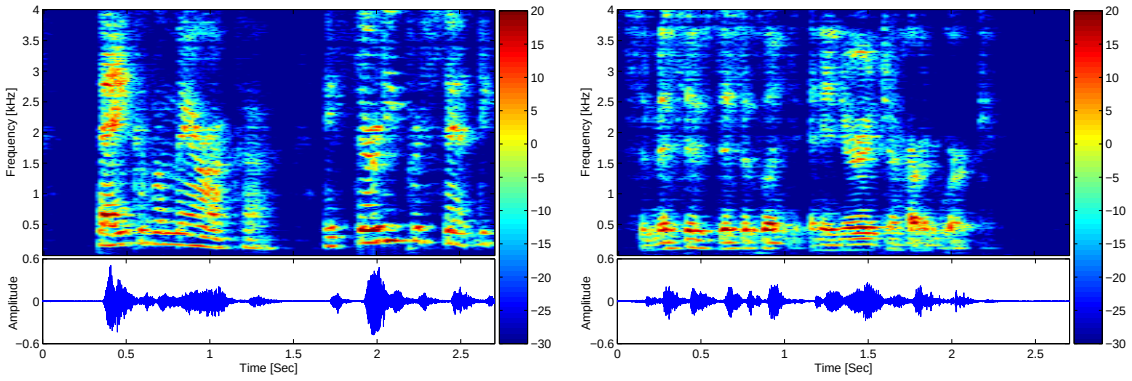
In this section, we have introduced the DGSC, a novel distributed algorithm for speech enhancement in multiple speakers, noisy and reverberant environment. It is proven analytically that the proposed algorithm converges to the optimal centralized GSC-BF. The adaptive procedure of the DGSC is based on the low complexity, time recursive NLMS algorithm. A common P linear constraints set, comprising the speakers' ATF, is shared by all nodes in the network. The algorithm requires $N+P$ transmission channels. The GSC structure splits the BF into two components. The first component lies in the constraints (speakers) subspace

and the second component lies in its corresponding null-space. The constraints subspace component of the DGSC is determined at the initialization phase of the algorithm where P shared signals are constructed by a selection procedure in the WASN. In static environments this procedure should be applied only at the initialization stage. The second component is implemented as an adaptive algorithm which converges in speech-absent time segments.

A comprehensive experimental study validates the equivalence between the centralized GSC and the DGSC algorithms. The proposed algorithm was tested successfully for both narrowband and speech signals in multiple Monte Carlo experiments.

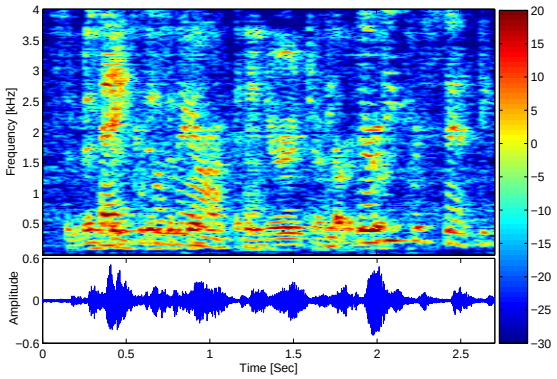
Table 3.2: Performance comparison of the centralized GSC, the DGSC and the single node GSC algorithms with speech signals.

Algorithm	SNR imp. [dB]	SIR imp. [dB]	Dist. [dB]
Cent. GSC	20.1	22.9	−23.0
DGSC	19.3	18.6	−20.6
1 node GSC	1.7	11.0	−14.1

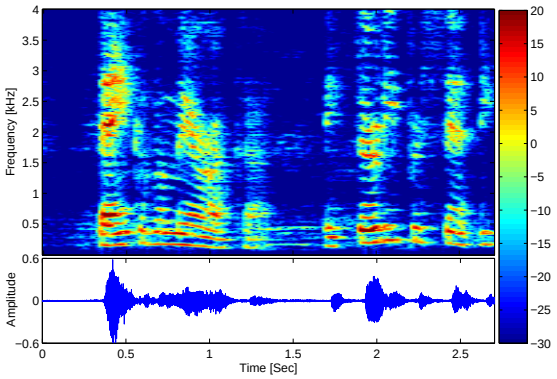


(a) Desired speaker.

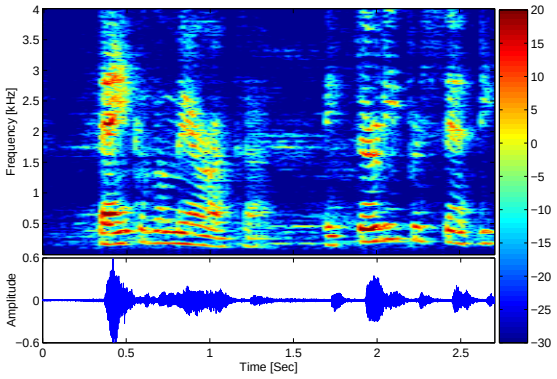
(b) Competing speaker.



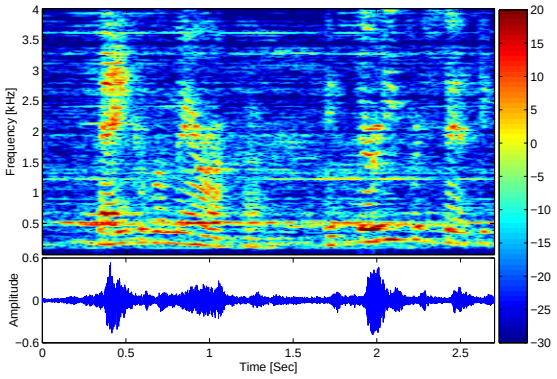
(c) First microphone.



(d) Centralized. GSC.



(e) DGSC.



(f) Single node GSC.

Figure 3.12: Sonograms of the various components of the signal received in the first microphone, and the outputs of the centralized GSC, the DGSC and the single node GSC.

3.3 Distributed GSC beamforming using the relative transfer function

A speech enhancement algorithm in a noisy and reverberant enclosure for a WASN is derived. We consider the case of a single desired speaker, which is a special case of the more general scenario dealt with in the previous Sec. 3.2. The algorithm derived in this section requires only N communication-channels, opposed to the DGSC (derived in Sec. 3.2) which requires $N+1$ channels in this case. The proposed algorithm is structured as a two stage BFs scheme, where the outputs of the first stage are broadcast in the network. Designing the second stage BF requires estimating the desired signal components at the transmitted signals. The contribution here is twofold. First, in spatially static scenarios, the first stage BFs are designed to maintain a fixed response towards the desired signal. As opposed to competing algorithms, where the response changes and repeated estimation thereof is required. Second, the proposed algorithm is implemented in a GSC form, separating the treatment of the desired speech and the interferences and enabling a simple time-recursive implementation of the algorithm. We outline a proof for the convergence of the proposed algorithm to the centralized TF-GSC. A comprehensive experimental study demonstrates the equivalent performance of the centralized GSC and of the proposed algorithm for both narrowband and speech signals.

The structure of the section is as follows. The problem is formulated in Sec. 3.3.1. In Sec. 3.3.2, the TF-GSC BF is surveyed. In Sec. 3.3.3, we introduce the proposed GSC-based distributed algorithm, namely the DS-GSC. A comprehensive experimental study is given in Sec. 3.3.4. Some concluding remarks are given in Sec. 3.3.5.

3.3.1 Problem formulation

Consider an N-node WASN. Denote by M_n the number of microphones in the nth node, and by $M = \sum_{n=1}^{N} M_n$ the total number of microphones. The problem is formulated in the STFT domain, where k denotes the frequency index and ℓ denotes the segment index. Let $\boldsymbol{z}(\ell,k)$ be a vector constructed by all received microphone signals $\boldsymbol{z}(\ell,k) = \left[\begin{array}{ccc} \boldsymbol{z}_1^T(\ell,k) & \cdots & \boldsymbol{z}_N^T(\ell,k) \end{array} \right]^T$ where $\boldsymbol{z}_n(\ell,k)$ is the local $M_n \times 1$ received signals vector of the nth node and $(\cdot)^T$ is the transpose operator. The term *local* (to the nth node) is associated with the signals and parameters calculated using only the microphones at the nth node. By the term *global* we

designate signals and parameters which are calculated using data from other nodes shared via the wireless link. The global vector of received signals is formulated as:

$$\boldsymbol{z}(\ell,k) = \boldsymbol{h}(\ell,k)s(\ell,k) + \boldsymbol{v}(\ell,k) \tag{3.99}$$

where $s(\ell,k)$ is the desired speech source, and $\boldsymbol{h}(\ell,k)$ consists of a vector of $M \times 1$ ATF from the desired speaker to the microphones. The vector $\boldsymbol{v}(\ell,k)$ is a vector of interferences picked up by microphones. Assuming that the speaker and the noise are statistically independent, the covariance matrix of the received signals is:

$$\boldsymbol{\Phi}_{zz}(\ell,k) = \lambda_s(\ell,k)\boldsymbol{h}(\ell,k)\boldsymbol{h}^H(\ell,k) + \boldsymbol{\Phi}_{vv}(\ell,k) \tag{3.100}$$

where $\lambda_s(\ell,k)$ is the variance of the desired speech signal, $\boldsymbol{\Phi}_{vv}(\ell,k)$ is the covariance matrix of the interferences and $(\cdot)^\dagger$ is the Hermitian operator. We assume that the network is fully connected, i.e. each transmitted signal is available to all nodes. In networks that are not fully connected a variation of the proposed algorithm can be derived, however, it is beyond the scope of this contribution. We assume that the location of the speaker is static and that the noise sources' statistics is slowly time-varying. Therefore, the speaker ATF and the noise covariance matrix are assumed time-invariant. Hence, in these quantities the frame index is omitted. For brevity, explicit frequency dependence is omitted hereafter. Finally, denote by $\boldsymbol{U}_n$ an $M \times M_n$ matrix that extracts the local microphones $\boldsymbol{z}_n(\ell) = \boldsymbol{U}_n^H \boldsymbol{z}(\ell)$ from $\boldsymbol{z}(\ell)$:

$$\boldsymbol{U}_n = \left[\begin{array}{c|c|c} \boldsymbol{0}_{M_n \times (\sum_{n'=1}^{n-1} M_{n'})} & \boldsymbol{I}_{M_n} & \boldsymbol{0}_{M_n \times (\sum_{n'=n+1}^{N} M_{n'})} \end{array} \right]^\dagger \tag{3.101}$$

where $\boldsymbol{I}_m$ is an $m \times m$ identity matrix.

The problem at hand is to enhance the desired speech signal at each node, with access only to the local microphones and the transmitted signals.

3.3.2 The centralized TF-GSC BF

Let $\boldsymbol{w}$ be the centralized TF-GSC BF. Recall that $\boldsymbol{w}$ is the output power, $\boldsymbol{w}^H\boldsymbol{\Phi}_{vv}\boldsymbol{w}$, minimizer, that satisfies the constraint $\boldsymbol{h}^H\boldsymbol{w} = 1$. In many applications, when the goal is to reduce the noise level, while sacrificing dereverberation, it is sufficient to enhance the desired signal as received by a reference microphone (arbitrarily chosen here to be the first microphone). The RTF [39], $\tilde{\boldsymbol{h}}$, is defined as the vector of ATFs from the desired signal to the

microphones normalized by the ATF to the reference microphone, $\tilde{\boldsymbol{h}} = \boldsymbol{h}/h_1$. The resulting modified constraint satisfies $\tilde{\boldsymbol{h}}^H \boldsymbol{w} = 1$ (see [39]).

Now, we are ready to formulate the centralized GSC:

$$\tilde{\boldsymbol{w}} = \tilde{\boldsymbol{q}} - \tilde{\boldsymbol{B}}\tilde{\boldsymbol{f}} \tag{3.102}$$

where $\tilde{\boldsymbol{q}}$ denotes the FBF and is parallel to the RTF, $\tilde{\boldsymbol{B}}$ denotes the BM and $\tilde{\boldsymbol{f}}$ denotes the NC. The FBF [25] is given by:

$$\tilde{\boldsymbol{q}} = \frac{\tilde{\boldsymbol{h}}}{\|\tilde{\boldsymbol{h}}\|^2}. \tag{3.103}$$

The matrix $\tilde{\boldsymbol{B}}$ is designed to block the RTF of the speaker, i.e., $\tilde{\boldsymbol{B}}^H \tilde{\boldsymbol{h}} = \boldsymbol{0}_{(M-1)\times 1}$, but it is not unique. One way of constructing the BM is by calculating the SVD of $\tilde{\boldsymbol{h}}$:

$$\tilde{\boldsymbol{h}} = \tilde{\boldsymbol{\Theta}}\tilde{\boldsymbol{\Gamma}}\tilde{\boldsymbol{\Psi}}^H \tag{3.104}$$

and selecting the $M-1$ left singular vectors (of $\tilde{\boldsymbol{\Theta}}$) which correspond to the zero singular values [25]. The NC is given in a closed-form by:

$$\tilde{\boldsymbol{f}} = \left(\tilde{\boldsymbol{B}}^H \boldsymbol{\Phi}_{vv} \tilde{\boldsymbol{B}}\right)^{-1} \tilde{\boldsymbol{B}}^H \boldsymbol{\Phi}_{vv} \tilde{\boldsymbol{q}}. \tag{3.105}$$

An efficient time-recursive implementation for adapting the NC during speech-absent segments utilizes the NLMS algorithm:

$$\tilde{\boldsymbol{f}}(\ell+1) = \tilde{\boldsymbol{f}}(\ell) + \frac{\mu}{\tilde{\lambda}_u(\ell)} \tilde{\boldsymbol{u}}(\ell)\tilde{y}^*(\ell) \tag{3.106}$$

where the noise reference signals at the output of the BM are denoted $\tilde{\boldsymbol{u}}(\ell) = \tilde{\boldsymbol{B}}^H \boldsymbol{z}(\ell)$, the output of the TF-GSC is denoted $\tilde{y}(\ell) = \tilde{\boldsymbol{w}}^H \boldsymbol{z}(\ell)$, the adaptation step is $0 < \mu < 2$, $\tilde{\lambda}_u(\ell)$ is a power normalization factor:

$$\tilde{\lambda}_u(\ell+1) = \rho\tilde{\lambda}_u(\ell) + (1-\rho)\|\tilde{\boldsymbol{u}}(\ell)\|^2, \tag{3.107}$$

and ρ is a forgetting factor (typically $0.8 < \rho < 1$).

3.3.3 The DS-GSC

A two-stage distributed enhancement TF-GSC BF algorithm, denoted as DS-GSC, is now proposed. The first stage consists of N TF-GSC-BFs which process local signals at each node. The output signals of the first stage are broadcasted in the WASN. The second stage consists of a global TF-GSC BF which processes the N outputs of the first stage. A replica of the global BF is concurrently applied in all nodes.

Denote the first stage output at the nth node as $\dot{y}_n(\ell)$ for $n = 1,..,N$. Let the concatenated N outputs of the first stage be $\dot{\boldsymbol{y}}_l(\ell) = \dot{\boldsymbol{W}}_l^H \boldsymbol{z}(\ell) = \left[\begin{array}{ccc} \dot{y}_1(\ell) \cdots & & \dot{y}_N(\ell) \end{array} \right]^T$, where the subscript l denotes *local*, the N local BFs are given in a matrix notation by $\dot{\boldsymbol{W}}_l = \left[\begin{array}{ccc} \boldsymbol{U}_1 \dot{\boldsymbol{w}}_1 & \cdots & \boldsymbol{U}_N \dot{\boldsymbol{w}}_N \end{array} \right]$, and $\dot{\boldsymbol{w}}_n$ are filters of the first stage TF-GSC BF at the nth node:

$$\dot{\boldsymbol{w}}_n = \dot{\boldsymbol{q}}_n - \dot{\boldsymbol{B}}_n \dot{\boldsymbol{f}}_n. \tag{3.108}$$

Note that $\dot{\boldsymbol{W}}_l$ is an $M \times N$ matrix. The local FBF, BM and NC at the nth node are denoted $\dot{\boldsymbol{q}}_n$, $\dot{\boldsymbol{B}}_n$ and $\dot{\boldsymbol{f}}_n$, respectively. The total output of the DS-GSC is given by $\dot{y}_g(\ell) = \dot{\boldsymbol{w}}_g^H \dot{\boldsymbol{y}}_l(\ell)$ where the subscript g denotes *global* and

$$\dot{\boldsymbol{w}}_g = \dot{\boldsymbol{q}}_g - \dot{\boldsymbol{B}}_g \dot{\boldsymbol{f}}_g \tag{3.109}$$

is the second stage global TF-GSC BF. The second stage FBF, BM and NC are denoted by $\dot{\boldsymbol{q}}_g$, $\dot{\boldsymbol{B}}_g$ and $\dot{\boldsymbol{f}}_g$, respectively. A block-diagram of the DS-GSC algorithm is depicted in Fig. 3.13.

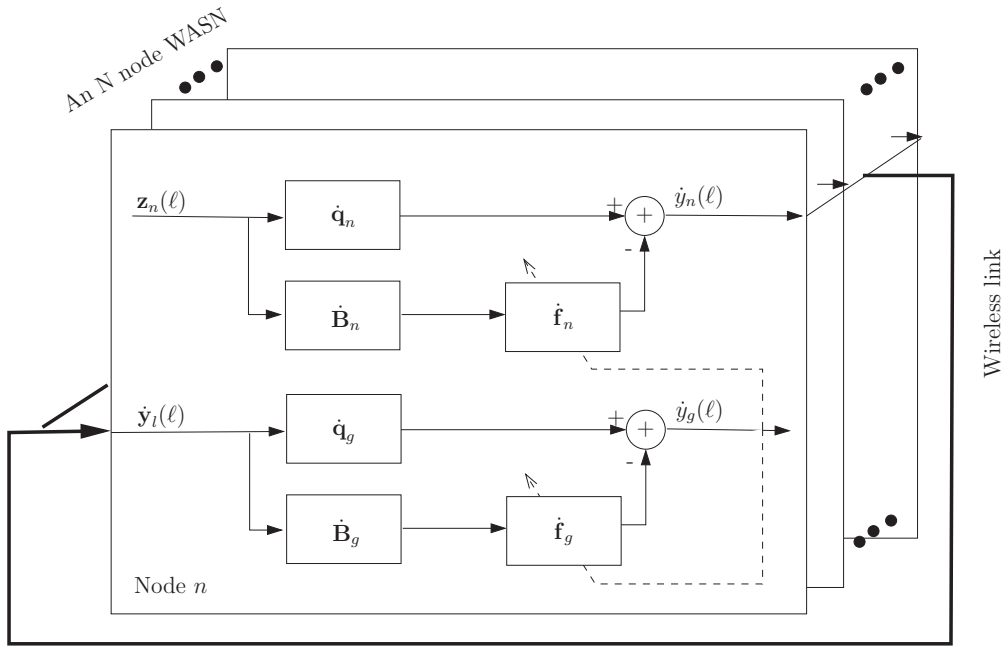


Figure 3.13: A block-diagram of the DS-GSC.

The proposed algorithm consists of three phases: first, the local RTFs are estimated, and the local FBFs and BMs are constructed; second, the global RTFs are estimated, and the global FBF and BM are constructed; third, the local and global NCs are alternatingly updated until convergence. The first two phases are only applied once in static environments. We adopt in our implementation a standard subspace-based RTF estimation procedure [51], where a perfect VAD is assumed.

The local and global stage filtering are presented in Secs. 3.3.3.1 and 3.3.3.2, respectively. In Sec. 3.3.3.3 we present an iterative algorithm for updating the NCs, and prove its convergence. Later, in Sec. 3.3.3.4 we derive a time-recursive variant.

3.3.3.1 Local stage filtering

Denote by $\dot{\boldsymbol{h}}_n$ the local RTF, which equals the ATF relating the desired signal and the local microphone signals at the nth node normalized by its first component, $\dot{\boldsymbol{h}}_n = \boldsymbol{U}_n^H \boldsymbol{h} / \left(\boldsymbol{U}_n^H \boldsymbol{h}\right)_1$. The local TF-GSC BF at the nth node is designed to satisfy the constraint $\dot{\boldsymbol{h}}_n^H \dot{\boldsymbol{w}}_n = 1$, therefore the desired signal component at its output is $h_{n,1} s(\ell)$. Similarly to (3.103) and

(3.104) the local FBF is given by:

$$\dot{\boldsymbol{q}}_n = \frac{\dot{\boldsymbol{h}}_n}{\|\dot{\boldsymbol{h}}_n\|^2} \tag{3.110}$$

and the local BM, $\dot{\boldsymbol{B}}_n$, is constructed by the $M_n - 1$ left singular vectors, corresponding to the zero singular values in the SVD of $\dot{\boldsymbol{h}}_n$.

3.3.3.2 Global stage filtering

Denote by $\dot{\boldsymbol{h}}_g$ the global RTF, which equals the ATF relating the desired signal and the local output signals $\dot{\boldsymbol{y}}_l(\ell)$, normalized by its first component $\dot{\boldsymbol{h}}_g = \frac{\dot{\boldsymbol{W}}_l^H \boldsymbol{h}}{h_1}$. Note, that the fixed response of the desired source at the local outputs, simplifies the global RTF estimation in static scenarios. Following similar arguments to (3.103) and (3.104) the global FBF is $\dot{\boldsymbol{q}}_g = \frac{\dot{\boldsymbol{h}}_g}{\|\dot{\boldsymbol{h}}_g\|^2}$ and the global BM, $\dot{\boldsymbol{B}}_g$, is constructed by the $N - 1$ left singular vectors, corresponding to the zero singular values in the SVD of $\dot{\boldsymbol{h}}_g$. Consequently, the desired signal component at the BF output is $h_1 s(\ell)$.

3.3.3.3 Iterative algorithm

Following the results of previous sections, the noise component at the output of the DS-GSC is:

$$\dot{v}(\ell) = \dot{\boldsymbol{w}}_g^H \dot{\boldsymbol{W}}_l^H \boldsymbol{v}(\ell) = \dot{\boldsymbol{w}}_l^H \dot{\boldsymbol{W}}_g^H \boldsymbol{v}(\ell). \tag{3.111}$$

where $\dot{\boldsymbol{W}}_g$ is an $M \times M$ diagonal matrix given by

$$\dot{\boldsymbol{W}}_g = \text{blkdiag}\left\{\dot{w}_{g,1}\boldsymbol{I}_{M_1}, \ldots, \dot{w}_{g,N}\boldsymbol{I}_{M_N}\right\} \tag{3.112}$$

and $\dot{\boldsymbol{w}}_l$ is a concatenation of the local GSC-BFs:

$$\dot{\boldsymbol{w}}_l = \dot{\boldsymbol{q}}_l - \dot{\boldsymbol{B}}_l \dot{\boldsymbol{f}}_l \tag{3.113}$$

and its components are given by:

$$\dot{\boldsymbol{q}}_l = \begin{bmatrix} \dot{\boldsymbol{q}}_1^T & \cdots & \dot{\boldsymbol{q}}_N^T \end{bmatrix}^T \tag{3.114a}$$

$$\dot{\boldsymbol{B}}_l = \text{blkdiag}\left\{\dot{\boldsymbol{B}}_1, .., \dot{\boldsymbol{B}}_N\right\} \tag{3.114b}$$

$$\dot{\boldsymbol{f}}_l = \begin{bmatrix} \dot{\boldsymbol{f}}_1^T & \cdots & \dot{\boldsymbol{f}}_N^T \end{bmatrix}^T. \tag{3.114c}$$

The variance of the noise component (3.111) is:

$$\gamma = \dot{\boldsymbol{w}}_g^H \dot{\boldsymbol{W}}_l^H \boldsymbol{\Phi}_{vv} \dot{\boldsymbol{W}}_l \dot{\boldsymbol{w}}_g \tag{3.115a}$$

$$= \dot{\boldsymbol{w}}_l^H \dot{\boldsymbol{W}}_g^H \boldsymbol{\Phi}_{vv} \dot{\boldsymbol{W}}_g \dot{\boldsymbol{w}}_l. \tag{3.115b}$$

Since the FBFs and BMs have been already determined (recall that the acoustic scenario is static), only the NC filters $\dot{\boldsymbol{f}}_l$ and $\dot{\boldsymbol{f}}_g$ should be set for minimizing the residual noise power. The concatenated NCs vector is denoted by $\dot{\boldsymbol{f}} = \begin{bmatrix} \dot{\boldsymbol{f}}_l^T & \dot{\boldsymbol{f}}_g^T \end{bmatrix}^T$.

We propose an iterative algorithm comprised of two alternating steps sequentially: first, updating the local NCs $\dot{\boldsymbol{f}}_l$; second, updating the global NC $\dot{\boldsymbol{f}}_g$. We denote values at the ith iteration with the superscript $(\cdot)^{(i)}$.

Consider the first update step, i.e., $\dot{\boldsymbol{f}}_g^{(i)}$ is updated to $\dot{\boldsymbol{f}}_g^{(i+1)}$ while $\dot{\boldsymbol{f}}_l^{(i+1)} = \dot{\boldsymbol{f}}_l^{(i)}$ remains fixed. An explicit expression of (3.115a) in terms of $\dot{\boldsymbol{f}}_g^{(i)}$ at the ith iteration is given by:

$$\gamma^{(i)} = \left(\dot{\boldsymbol{q}}_g - \dot{\boldsymbol{B}}_g \dot{\boldsymbol{f}}_g^{(i)}\right)^{\dagger} \left(\dot{\boldsymbol{W}}_l^{(i)}\right)^H \boldsymbol{\Phi}_{vv} \dot{\boldsymbol{W}}_l^{(i)} \left(\dot{\boldsymbol{q}}_g - \dot{\boldsymbol{B}}_g \dot{\boldsymbol{f}}_g^{(i)}\right). \tag{3.116}$$

Equation (3.116) is a quadratic function of $\dot{\boldsymbol{f}}_g^{(i)}$, allowing for a simple calculation of the gradient with respect to $\dot{\boldsymbol{f}}_g$:

$$\frac{\partial \gamma^{(i)}}{\partial \dot{\boldsymbol{f}}_g^H} = -\dot{\boldsymbol{B}}_g^H \left(\dot{\boldsymbol{W}}_l^{(i)}\right)^H \boldsymbol{\Phi}_{vv} \dot{\boldsymbol{W}}_l^{(i)} \left(\dot{\boldsymbol{q}}_g - \dot{\boldsymbol{B}}_g \dot{\boldsymbol{f}}_g\right). \tag{3.117}$$

The updated $\dot{\boldsymbol{f}}_g$ is obtained by equating the gradient to zero:

$$\dot{\boldsymbol{f}}_g^{(i+1)} = \left(\dot{\boldsymbol{B}}_g^H \left(\dot{\boldsymbol{W}}_l^{(i)}\right)^H \boldsymbol{\Phi}_{vv} \dot{\boldsymbol{W}}_l^{(i)} \dot{\boldsymbol{B}}_g\right)^{-1} \cdot \dot{\boldsymbol{B}}_g^H \left(\dot{\boldsymbol{W}}_l^{(i)}\right)^H \boldsymbol{\Phi}_{vv} \dot{\boldsymbol{W}}_l^{(i)} \dot{\boldsymbol{q}}_g. \tag{3.118}$$

Consider the second update step, i.e. $\dot{\boldsymbol{f}}_l^{(i)}$ is updated to $\dot{\boldsymbol{f}}_l^{(i+1)}$ while $\dot{\boldsymbol{f}}_g^{(i+1)} = \dot{\boldsymbol{f}}_g^{(i)}$ remains fixed. An explicit expression of (3.115b) at the ith iteration is given in terms of $\dot{\boldsymbol{f}}_l$ by:

$$\gamma^{(i)} = \left(\dot{\boldsymbol{q}}_l - \dot{\boldsymbol{B}}_l \dot{\boldsymbol{f}}_l^{(i)}\right)^{\dagger} \left(\dot{\boldsymbol{W}}_g^{(i)}\right)^H \dot{\boldsymbol{\Phi}}_{vv} \dot{\boldsymbol{W}}_g^{(i)} \left(\dot{\boldsymbol{q}}_l - \dot{\boldsymbol{B}}_l \dot{\boldsymbol{f}}_l^{(i)}\right). \tag{3.119}$$

Equation (3.119) is a quadratic function of $\dot{\boldsymbol{f}}_l$. The gradient of $\gamma^{(i)}$ with respect to $\dot{\boldsymbol{f}}_l$ is:

$$\frac{\partial \gamma^{(i)}}{\partial \dot{\boldsymbol{f}}_l^H} = -\dot{\boldsymbol{B}}_l^H \left(\dot{\boldsymbol{W}}_g^{(i)}\right)^H \dot{\boldsymbol{\Phi}}_{vv} \dot{\boldsymbol{W}}_g^{(i)} \left(\dot{\boldsymbol{q}}_l - \dot{\boldsymbol{B}}_l \dot{\boldsymbol{f}}_l\right) \tag{3.120}$$

yielding

$$\begin{aligned} \dot{\boldsymbol{f}}_l^{(i+1)} = &\left(\dot{\boldsymbol{B}}_l^H \left(\dot{\boldsymbol{W}}_g^{(i)}\right)^H \dot{\boldsymbol{\Phi}}_{vv} \dot{\boldsymbol{W}}_g^{(i)} \dot{\boldsymbol{B}}_l\right)^{-1} \\ &\cdot \dot{\boldsymbol{B}}_l^H \left(\dot{\boldsymbol{W}}_g^{(i)}\right)^H \dot{\boldsymbol{\Phi}}_{vv} \dot{\boldsymbol{W}}_g^{(i)} \dot{\boldsymbol{q}}_l. \end{aligned} \tag{3.121}$$

It can be easily verified that the proposed algorithm converges. First, $\gamma^{(i+1)} \leq \gamma^{(i)}$ is monotonically non-increasing, since $\dot{\boldsymbol{f}}^{(i)}$ belongs to the minimization range of $\dot{\boldsymbol{f}}^{(i+1)}$. Second, $\gamma^{(i)}$ is trivially lower bounded by $0 \leq \gamma^{(i)}$. In practice the iterative algorithm cannot be implemented, since updating $\dot{\boldsymbol{f}}_l$ (3.121) involves non-local quantities unavailable at each node. However, a practical time-recursive algorithm can be derived, as presented in the sequel.

3.3.3.4 Time-recursive algorithm

A time-recursive procedure for updating the NCs is obtained by using two NLMS algorithms for updating $\dot{\boldsymbol{f}}_g(\ell)$ and $\dot{\boldsymbol{f}}_l(\ell)$ alternately, during speech-absent time segments. As in all stochastic approximation procedures, we propose to replace the iteration index i by the time-segment index ℓ. We further propose to perform L_u updates of each filter before switching to the other filter updates.

Consider an update step of $\dot{\boldsymbol{f}}_g(\ell)$ to $\dot{\boldsymbol{f}}_g(\ell+1)$ while $\dot{\boldsymbol{f}}_l(\ell+1) = \dot{\boldsymbol{f}}_l(\ell)$ is unaltered. An instantaneous estimate of the gradient (3.117) at the ℓth frame is $-\dot{\boldsymbol{u}}_g(\ell)\dot{v}^*(\ell)$ where $\dot{\boldsymbol{u}}_g(\ell) = \dot{\boldsymbol{B}}_g^H \dot{\boldsymbol{v}}_l(\ell)$, and $\dot{\boldsymbol{v}}_l(\ell)$ is the noise component in $\dot{\boldsymbol{y}}_l(\ell)$. The updated $\dot{\boldsymbol{f}}_g(\ell+1)$ is therefore:

$$\dot{\boldsymbol{f}}_g(\ell+1) = \dot{\boldsymbol{f}}_g(\ell) + \frac{\mu}{\dot{\lambda}_{u,g}(\ell)} \dot{\boldsymbol{u}}_g(\ell)\dot{v}^*(\ell) \tag{3.122}$$

where, similarly to (3.107), $\dot{\lambda}_{u,g}(\ell)$ is a time-recursive estimate of the power normalization. Similarly, we can define an update step of $\dot{\boldsymbol{f}}_l(\ell)$ to $\dot{\boldsymbol{f}}_l(\ell+1)$ while $\dot{\boldsymbol{f}}_g(\ell+1) = \dot{\boldsymbol{f}}_g(\ell)$. An instantaneous estimate of the gradient (3.120) at the ℓth frame is $-\dot{\boldsymbol{W}}_g^H(\ell)\dot{\boldsymbol{u}}_l(\ell)\dot{v}^*(\ell)$, where $\dot{\boldsymbol{u}}_l(\ell) = \dot{\boldsymbol{B}}_l^H\boldsymbol{z}(\ell)$. Note that the calculation of the nth component of the estimated gradient can be done locally. Hence, $\dot{\boldsymbol{f}}_n(\ell+1);\ n = 1,..,N$ can be updated in parallel:

$$\dot{\boldsymbol{f}}_n(\ell+1) = \dot{\boldsymbol{f}}_n(\ell) + \frac{\mu}{\dot{\lambda}_{u,l}^n(\ell)}\dot{w}_{g,n}^*(\ell)\dot{\boldsymbol{u}}_n(\ell)\dot{v}^*(\ell) \tag{3.123}$$

where, as in (3.107), $\dot{\lambda}_{u,l}^n(\ell)$ is a power normalization factor.

3.3.4 Experimental study

An experimental study of three algorithms using multiple Monte-Carlo trials is presented. We compare the centralized TF-GSC, the time-recursive DS-GSC and the local TF-GSC (a TF-GSC BF which utilizes only the microphones of a the first node). A WASN comprised of $N = 4$ nodes, each consisting of $M_n = 2$ microphones was simulated. The performance of the BFs was evaluated by using the SIR improvement.

3.3.4.1 Narrowband signals

The performance of the algorithms was tested with various numbers of interfering sources $(1, 3, 5, 7)$. For each number of interfering sources, 20 different sets of desired source ATFs and noise covariance matrices were randomized at a particular frequency bin. For each set, 10 realizations of 10^5 samples of the desired and interfering narrowband sources were randomized (white Gaussian processes). Also, a spatially white noise was added to the received signals. The input SIR and SNR were set to 5dB and 30dB, respectively. In total, 800 Monte-Carlo experiments were tested. The algorithm parameters were set to: $\mu = 0.15$, $\rho = 0.95$ and $L_u = 12$. The SIR improvement of the proposed DS-GSC exhibits equivalent performance to the centralized TF-GSC in all Monte-Carlo experiments, and both significantly outperform the local TF-GSC. The average SIR improvement versus the number of interferences is depicted in Fig. 3.14.

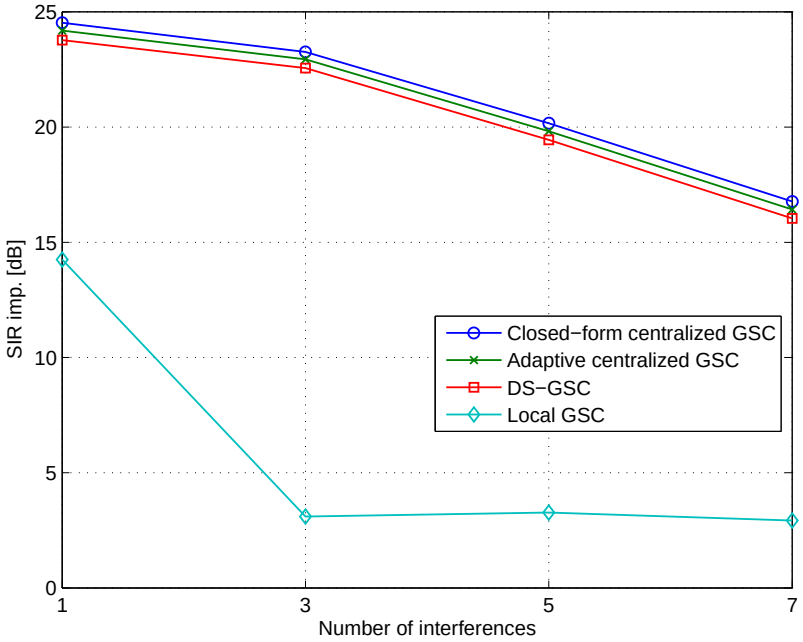


Figure 3.14: SIR improvement versus the number of interferences with narrowband signals.

3.3.4.2 Speech signals

A 4m × 3m × 3m room with a reverberation time of $T_{60} = 300$ms was simulated. Four nodes were arbitrarily positioned at the center of the walls. The locations of the desired speaker and the two interfering sources were randomized with a minimum distance of 10cm from the walls in 100 Monte-Carlo experiments. An additive spatially white noise was added to the received signals. The input SIR and SNR were set to 6dB and 60dB, respectively. The sample rate was 8kHz and a 4096 point STFT with 75% overlap was used. The algorithm parameters were set to $\mu = 0.15$, $\rho = 0.95$ and $L_u = 12$. The average SIR improvement of the centralized TF-GSC, the DS-GSC and the local TF-GSC were 26.4dB, 28.6dB and 12.2dB, respectively. Theoretically, the DS-GSC cannot outperform the centralized TF-GSC, but in our experiments a 2.2dB improvement of the DS-GSC was recorded. This can be attributed to the finite segment length and the lengths of the different equivalent filters applied in each scheme.

3.3.5 Conclusions

A DS-GSC algorithm was proposed and both iterative and time-recursive adaptation procedures were derived. In static scenarios, only a single estimate of the speech RTF is required, as opposed to closed-form based distributed algorithms, which require repeated RTFs estimations. The convergence of the proposed algorithm was proven. A comprehensive experimental study demonstrates the equivalence between the centralized TF-GSC and the proposed DS-GSC.

3.4 Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming

In this section, we address the problem of blind sampling rate offset estimation and compensation in beamforming applications. As we are interested in blind scenarios, no reference signals (neither speech from a far-end nor pilot channels) are assumed to be available. To simplify the exposition, a WASN, for which all microphone signals are available at the fusion center, is considered. The same sampling rate compensation method is applicable in distributed constellations as well. The proposed method utilizes speech-absent time segments, where the interference statistics is assumed slowly time-varying, and the sampling rate offsets are assumed fixed [77]. An estimation procedure for the sampling rate offsets is proposed based on the coherence between the received signals. Following [80], we propose to resample the microphone signals with the Lagrange polynomials interpolation method. We incorporate the proposed sampling rate offset estimation and compensation scheme in the TF-GSC, introduced by Gannot et al. [39].

The structure of the section is as follows. In Sec. 3.4.1, the problem is defined. An algorithm for estimating the sampling rate offsets is derived in Sec. 3.4.2. The Lagrange polynomials interpolation method is described in Sec. 3.4.3. In Sec. 3.4.4, the sampling rate offset estimation and compensation procedures are incorporated in the TF-GSC-BF. The performance of the proposed structure is evaluated in an extensive simulated experimental study. Conclusions are drawn in Sec. 3.4.5.

3.4.1 Problem formulation

Consider a WASN comprising N nodes in a reverberant and noisy enclosure. The nth node comprises M_n microphones, and the total number of microphones is $M = \sum_{n=1}^{N} M_n$. Denote the $M_n \times 1$ microphone signals vector in continuous time at the nth node by:

$$\boldsymbol{z}_n(t) = \boldsymbol{x}_n(t) + \boldsymbol{v}_n(t). \tag{3.124}$$

Let $z_{n,r}(t)$ be the continuous-time received signal at microphone r in node n, and $x_{n,r}(t)$, $v_{n,r}(t)$ the respective speech and noise components. It is assumed that the received noise contains a spatially coherent component, and that a perfect VAD is available. Hence, noise-only segments can be determined, and hereafter only these segments will be considered.

Since nodes utilize individual clock sources, sampling rate differences are inevitable. Denote the sample rate at the nth node by $f_{s,n}$. Without loss of generality, the sampling rate offsets are defined with respect to the first node by:

$$f_{s,n} = (1 + \epsilon_n) f_s \tag{3.125}$$

where ϵ_n is the relative sampling rate offset and $f_s = f_{s,1}$ is the sampling rate at the first node.

The sampled microphone signals at the nth node are denoted by $\boldsymbol{v}_n[p] = \boldsymbol{v}_n(pT_{s,n})$ where $T_{s,n} = \frac{1}{f_{s,n}}$ is the sampling period at the nth node and p is the sample index. The notations $(\bullet)$ and $[\bullet]$ are used for denoting continuous-time and discrete-time functions, respectively. We assume that the WASN is fully connected and that all sampled microphone signals are transmitted to the fusion center, selected, without loss of generality, as the first node.

Let $V_{n,r}^{\ell}[k]$ be the discrete STFT of the rth microphone signal of the nth node at the ℓth sample, using the analysis window $c[p]$:

$$V_{n,r}^{\ell}[k] = \sum_{p=-\infty}^{\infty} v_{n,r}[p]\, c[p-\ell] \exp\left(-\mathrm{j}\frac{2\pi kp}{K}\right) \tag{3.126}$$

where $k = 0, 1, \ldots, K-1$. Throughout the section, lowercase and uppercase letters denote functions in time and frequency domains, respectively.

Since $\boldsymbol{v}_n[p]\,;\ n = 1, \ldots, N$ are sampled with different sampling rates, straightforward application of a beamforming algorithm may result in a degraded performance. In Sec. 3.4.4

we will demonstrate this degradation. In the next section, a sampling rate offset estimation procedure is derived.

3.4.2 Sampling rate offset estimation

We turn now to the derivation of a procedure for estimating ϵ_n, the sampling rate offset of the nth node. In Sec. 3.4.2.1 some notations are defined, and in Sec. 3.4.2.2 the procedure is derived.

3.4.2.1 Notation

Consider microphones s and r at the first and the nth nodes, respectively. Let $R_{s,r}(\tau)$ and $\theta_{s,r}(\zeta)$ be their cross-correlation and cross-spectrum:

$$R_{s,r}(\tau) = \mathrm{E}\{v_{1,s}(t)\, v_{n,r}(t-\tau)\} \tag{3.127a}$$

$$\theta_{s,r}(\zeta) = \int_{-\infty}^{\infty} R_{s,r}(\tau) \exp(-\mathrm{j}\zeta\tau)\, \mathrm{d}\tau \tag{3.127b}$$

where we assume that $v_{1,s}(t)$ and $v_{n,r}(t)$ are jointly wide-sense stationary (WSS) processes. Similarly to (3.127a), $R_{s,r}(\tau+\Delta) = \mathrm{E}\{v_{1,s}(t)\, v_{n,r}(t-\tau-\Delta)\}$, and the corresponding cross-spectrum, denoted $\theta^{\Delta}_{s,r}(\zeta)$, is obtained by applying Fourier transform properties:

$$\theta^{\Delta}_{s,r}(\zeta) = \exp(\mathrm{j}\zeta\Delta)\, \theta_{s,r}(\zeta). \tag{3.128}$$

Due to sampling rate offset, the time difference between the ℓth sample at microphone s and microphone r is approximately $\ell T_s - \ell T_{s,n} \approx \ell T_s \epsilon_n$, where we replaced $T_{s,n} = \frac{T_s}{1+\epsilon_n}$ with its first-order Taylor series approximation $T_s(1-\epsilon_n)$. Let $\theta^{\ell}_{s,r}[k]$ be the discrete cross-spectrum of the sampled microphones s and r at the ℓth sample. Assuming that the continuous cross-spectrum is band-limited by $\frac{f_s}{2}$, the following equivalence between the discrete and the continuous spectra holds:

$$\theta^{\ell}_{s,r}[k] = \theta^{\ell T_s \epsilon_n}_{s,r}\left(\frac{2\pi k f_s}{K}\right). \tag{3.129}$$

Now, applying the relation in (3.128) to (3.129) yields:

$$\theta^{\ell}_{s,r}[k] = \exp\left(\mathrm{j}\frac{2\pi k \ell \epsilon_n}{K}\right) \theta_{s,r}[k] \tag{3.130}$$

where

$$\theta_{s,r}[k] = \theta_{s,r}\left(\frac{2\pi k f_s}{K}\right). \tag{3.131}$$

Let $\theta_{s,s}[k]$ and $\theta_{r,r}[k]$ be the auto-spectra of microphones s and r, respectively. Denote the coherence between microphones s and r at the ℓth sample by:

$$\gamma^{\ell}_{s,r}[k] = \frac{\theta^{\ell}_{s,r}[k]}{\sqrt{\theta_{s,s}[k]\theta_{r,r}[k]}} \tag{3.132}$$

for $k = 0, 1, \ldots, K-1$. Substituting (3.130) in (3.132), the coherence is given by:

$$\gamma^{\ell}_{s,r}[k] = \alpha^{\ell}_{n}\gamma_{s,r}[k] \tag{3.133}$$

where

$$\gamma_{s,r}[k] = \frac{\theta_{s,r}[k]}{\sqrt{\theta_{s,s}[k]\theta_{r,r}[k]}} \tag{3.134a}$$

$$\alpha_n = \exp\left(\mathrm{j}\frac{2\pi k \epsilon_n}{K}\right). \tag{3.134b}$$

3.4.2.2 Estimation

We propose the following procedure for estimating ϵ_n; $n = 2, \ldots, N$, given P_s samples of the microphone signals. We wish to estimate the coherence between the microphones at the first and the nth node, $n = 2, \ldots, N$, at samples $\ell = i \times P$; $i = 0, 1, \ldots, I-1$, where $I = \lfloor \frac{P_s}{P} \rfloor$. For that purpose the Welch method with K frequency bins is applied. Note that the support of the analysis window $c[n]$ should be significantly longer than the delay induced by the sampling rate offset in order the accurately estimate ϵ_n. The estimated $M_1 \times M_n$ cross-coherence matrix at sample iP is denoted by $\hat{\mathbf{\Gamma}}^{iP}_{1,n}[k]$. Assuming that the estimation error is low $\hat{\gamma}^{iP}_{s,r}[k] \approx \gamma^{iP}_{s,r}[k]$, where $\hat{\gamma}^{iP}_{s,r}[k]$ is the element in the sth row and the rth column of the matrix $\hat{\mathbf{\Gamma}}^{iP}_{1,n}[k]$. Considering (3.133) and assuming that the frequency offsets between the transformed microphone signals are significantly smaller than the frequency resolution, we note that $\alpha^{P}_{n} = \frac{\gamma^{iP}_{s,r}[k]}{\gamma^{(i-1)P}_{s,r}[k]}$.

Assume that the sampling rate offset is bounded to $|\epsilon_n| < \epsilon_{\max}$, and define

$$k_{\max} = \frac{K}{2P\epsilon_{\max}}. \tag{3.135}$$

Note that for $1 \leq k \leq k_{\max}$ it is guaranteed that the phase difference between $\gamma_{s,r}^{iP}[k]$ and $\gamma_{s,r}^{(i-1)P}[k]$ is bounded in the range $[-\pi, \pi]$. Therefore, we propose to estimate ϵ_n by averaging the results obtained from all microphone couples:

$$\hat{\epsilon}_n = \frac{1}{M_1 M_n} \sum_{s=1}^{M_1} \sum_{r=1}^{M_n} \hat{\epsilon}_{n,s,r} \tag{3.136}$$

where $\hat{\epsilon}_{n,s,r}$ is the estimate of the sampling rate offset derived from microphones s and r. It is obtained by averaging the phase differences of consecutive estimates of $\hat{\gamma}_{s,r}^{iP}[k]$ and $\hat{\gamma}_{s,r}^{(i-1)P}[k]$ for all k in the allowable range with a proper frequency dependent normalization factor:

$$\hat{\epsilon}_{n,s,r} = \frac{1}{k_{\max}} \sum_{k=1}^{k_{\max}} \frac{K}{2\pi P k} \angle \left\{ \frac{1}{I-1} \left(\sum_{i=1}^{I-1} \frac{\hat{\gamma}_{s,r}^{iP}[k]}{\gamma_{s,r}^{(i-1)P}[k]} \right) \right\}. \tag{3.137}$$

Note that the averaging is applied in both time and frequency. If $|\epsilon_n| > \epsilon_{\max}$, a phase ambiguity of 2π will occur and the estimated offset will be incorrect.

In the following section, given estimates of $\epsilon_n;\ n = 2, \ldots, N$, we describe a procedure, applied by the fusion center, that compensates for sampling rate offsets by resampling the microphone signals.

3.4.3 Resampling with Lagrange polynomials interpolation

Consider the rth microphone signal of the nth node, i.e. $z_{n,r}[p]$. Given an estimate of the sampling rate offset at the nth node, $\hat{\epsilon}_n$, the signal is resampled to the sampling rate of the fusion center, f_s, by a fourth order Lagrange polynomials interpolation [81]. First, $z_{n,r}[p]$ is interpolated by a factor of 4, and the signal $\tilde{z}_{n,r}[\tilde{p}]$ is obtained. Denote $\dot{p} = \lfloor \frac{4pT_s}{T_{s,n}} \rfloor \approx \lfloor 4p(1+\hat{\epsilon}_n) \rfloor$, the closest interpolated sample index preceding time pT_s. Then, the resampled value of $z_{n,r}(pT_s)$, denoted $\hat{z}_{n,r}[p]$, is calculated by proper weighting its four neighboring interpolated samples:

$$\begin{aligned} \hat{z}_{n,r}[p] = &\beta_{-1}^p \tilde{z}_{n,r}[\dot{p}-1] + \beta_0^p \tilde{z}_{n,r}[\dot{p}] \\ &+ \beta_1^p \tilde{z}_{n,r}[\dot{p}+1] + \beta_2^p \tilde{z}_{n,r}[\dot{p}+2] \end{aligned} \tag{3.138}$$

where

$$\eta = 4p\left(1+\hat{\epsilon}_n\right) - \dot{p} \tag{3.139a}$$

$$\beta_{-1}^{p} = -\frac{\eta\left(\eta-1\right)\left(\eta-2\right)}{6} \tag{3.139b}$$

$$\beta_{0}^{p} = \frac{\left(\eta+1\right)\left(\eta-1\right)\left(\eta-2\right)}{2} \tag{3.139c}$$

$$\beta_{1}^{p} = -\frac{\left(\eta+1\right)\eta\left(\eta-2\right)}{2} \tag{3.139d}$$

$$\beta_{2}^{p} = \frac{\left(\eta+1\right)\eta\left(\eta-1\right)}{6}. \tag{3.139e}$$

3.4.4 Experimental study

Consider the following scenario. A $4\text{m}\times3\text{m}\times3\text{m}$ room, with a reverberation time of 300ms is simulated. A desired speaker and Q point source stationary interfering sources are picked up by the microphones, for $Q = 1,\ldots,4$. Utterances of 75s with 20% voice activity and a 6dB SNR are used. Two microphone arrays, each comprises 6 microphones with 5cm spacing, are located close to two perpendicular walls. The sampling rate of the first array is set to 8kHz, whereas the sampling rate of the second array is subject to offsets in the range of $\{-300,-250,\ldots,300\}$ ppm of the sampling rate of the first array, where ppm $= 10^{-6}$. The sampling rate offsets are simulated using the Lagrange polynomials interpolation method, discussed above. For each combination of sampling rate offset and number of interferences, 5 Monte-Carlo experiments are conducted, where the locations of the sources are randomly selected. The proposed sampling rate estimation and compensation scheme incorporated in the TF-GSC [39] is denoted the synchronized TF-GSC. The performances of the regular TF-GSC and the synchronized TF-GSC are compared in the various scenarios. The RTF is estimated once, using the subspace method [51], and is used to construct the FBF and the BM, which remain fixed during the entire utterance. The NC is adapted using the NLMS. The performance criteria are the excess distortion and the excess noise with respect to the corresponding TF-GSC without a sampling rate offset. The following parameters are used in the proposed sampling rate estimation. The Welch method with a discrete Fourier transform (DFT) size of 4096, 75% overlap and a Hamming window is applied to 32s speech-absent segments for estimating the auto and cross-covariances $\theta_{s,s}\left[k\right]$, $\theta_{r,r}\left[k\right]$ and $\theta_{s,r}^{\ell}\left[k\right]$. Coherence estimates, $\hat{\mathbf{\Gamma}}_{1,n}^{iP}\left[k\right]$, of $I = 6$ time segments with 50% overlap ($P = 128\times10^3$), are used for estimating the sampling rate offset, assuming that it is bounded by $|\epsilon_n| \leq 400$ppm.

The TF-GSC uses a 4096 points STFT with 75% overlap, and an NLMS step of $\mu = 0.15$.

The results of the experimental study, averaged over all frequency offsets. are summarized in Table. 3.3. The standard deviation of the estimated sampling rate offset in the synchronized TF-GSC is lower than 3.2ppm in all scenarios.

Table 3.3: Performance (excess noise and distortion) of the regular and the synchronized TF-GSC with sampling rate offsets

	Without offset		With offset			
	Regular		Regular		Synchronized	
Q	SDR	SNR	Ex. Dist.	Ex. Noise	Ex. Dist.	Ex. Noise
1	15.0	34.3	11.2	7.7	0.0	0.0
2	14.9	27.5	11.2	4.9	0.1	0.0
3	14.6	24.5	11.5	3.4	0.4	0.1
4	14.7	23.5	11.9	2.9	0.8	0.2

Clearly, the performance of the proposed synchronized TF-GSC is comparable to the regular TF-GSC without sampling rate offsets and is highly superior to the regular TF-GSC (with sampling rate offsets). The excess noise level in the regular TF-GSC with respect to its counterpart without a sampling rate offset is depicted in Fig. 3.15.

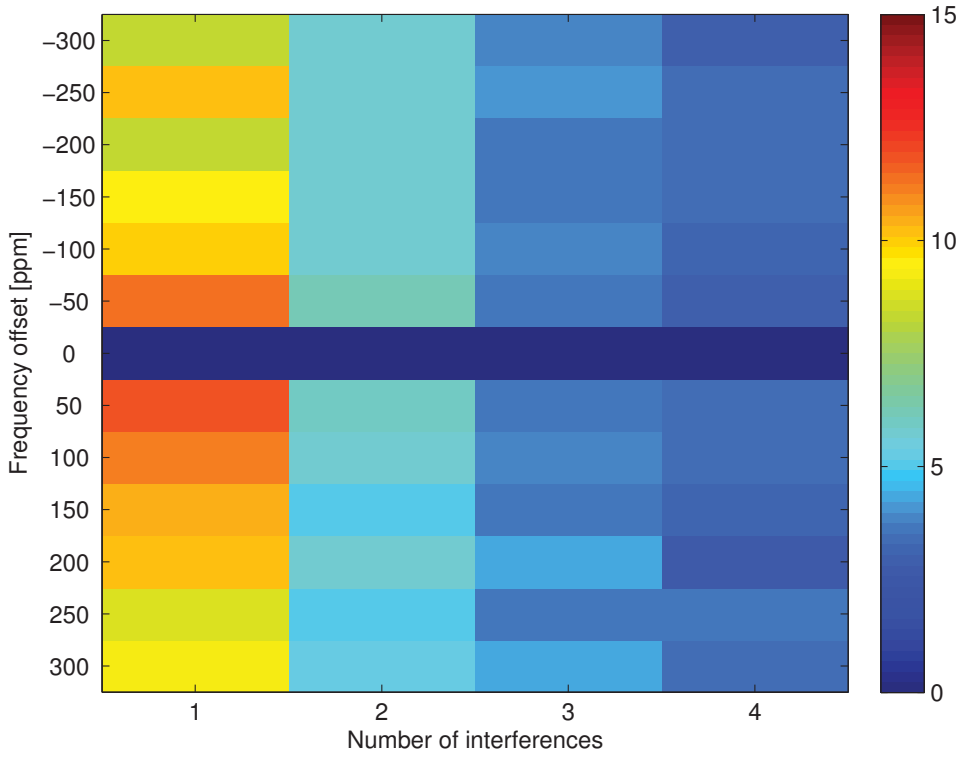


Figure 3.15: The excess noise (dB) in the regular TF-GSC with respect to its counterpart without a sampling rate offset.

3.4.5 Conclusions

The problem of blind sampling rate offset estimation and compensation in WASN was considered. An interference with slowly time-varying statistics, and speech-absent segments obtained by a perfect VAD were assumed available. A procedure for estimating the sampling rate offsets was derived. It was based on the phase drift in the coherence between two microphone signals, sampled at different sampling rates. The estimated sampling rate offsets were compensated for by resampling the signals with the Lagrange polynomials interpolation method. The estimation and compensation scheme was incorporated in the TF-GSC, denoted the synchronized TF-GSC. It was shown that, under sampling rate offsets, the synchronized TF-GSC significantly outperforms the regular TF-GSC. Moreover, the performance of the proposed method is comparable to the performance of the regular TF-GSC as if there were no sampling rate offsets.

Chapter 4

Performance analysis of arrays with randomly located microphones

Careful design of the microphone array is usually possible in condensed arrays. However, in some WASNs the exact nodes/microphones positions cannot be controlled. As the beamforming performance depends on microphone locations, a classical deterministic analysis does not suit this problem. Instead, following Lo [6], we treat the microphone locations as stochastic variables, and derive probabilistic measures of the performance. Note that previous contributions in the literature considered only data-independent BFs in non-reverberant environments.

In this chapter the performance of data-dependent BFs is analyzed for scenarios where the microphones are randomly located. Since, the performance depends on the microphones' constellation, it is also a random variable (RV). We analyze its statistics in two scenarios. The derived statistics enables the system designer to determine the number of microphones required to meet a desired performance level (with a controlled uncertainty).

In Sec. 4.1, a scenario of microphones uniformly distributed over a line is considered. A single desired speaker, and a single noise source arriving from the far-field in a non-reverberant environment are picked up by the microphone array. We analyze the statistics of the obtained performance, measured by the noise reduction level, of an MWF-BF.

In Sec. 4.2, we extend the work of Sec. 4.1 and consider a reverberant enclosure with randomly located microphones with an arbitrary constellation. A desired speaker contaminated by two types of noise sources is considered: 1) P coherent noise sources; 2) a diffuse noise field. The performance of the SDW-MWF, measured as the noise-reduction and sensitivity levels, is analyzed for these cases. We derive a statistical model for the performance, and

verify it in a simulative experimental study.

4.1 Performance analysis of a randomly spaced wireless microphone array

In this section a statistical model for the performance of the MWF beamformer is derived for a randomly located linear microphone array in a typical speech enhancement scenario. As stated earlier, the MWF is considered here since data-dependent beamformers are more suitable to speech processing than their data-independent counterparts [29]. We treat the scenario of a coherent wideband desired source and a coherent wideband interfering source arriving from the far-field and impinging on the microphone array, in a non-reverberant environment. Other scenarios can be treated in a similar fashion.

In Sec. 4.1.1, the problem is formulated. In Sec. 4.1.2, a formula for the MMSE of the MWF, given the microphones locations, is derived. Then, in Secs. 4.1.3 and 4.1.4 the statistics of the MMSE cost function is analyzed. The derived theoretical models are verified in Sec. 4.1.5. Aspects of applying the MWF in randomly distributed microphone arrays to speech processing are discussed in Sec. 4.1.6.

4.1.1 Problem Formulation

Consider a coherent wideband desired source and a coherent wideband interfering source impinging on a linear array of randomly spaced microphones from the far-field in a reverberant-free environment. The array is assumed to comprise M uniformly distributed microphones in the range $\left[-\frac{\Delta}{2}, \frac{\Delta}{2}\right]$, where Δ is the array aperture. The microphones locations are denoted by $x_1, \ldots, x_M$. In the STFT domain, the desired source is denoted by $s_d(\ell, k)$, and the interfering source is denoted by $s_i(\ell, k)$, where ℓ is the frame index, and k is the frequency index. The analysis window length is denoted by N_{DFT}. For simplicity, the desired source is assumed to arrive from the broadside. The angle between the interfering and the desired sources is denoted by θ_i. The received signals are denoted in vector notation by:

$$\boldsymbol{z}(\ell, k) = \sqrt{M}\boldsymbol{h}_d(\ell, k)s_d(\ell, k) + \sqrt{M}\boldsymbol{h}_i(\ell, k)s_i(\ell, k) + \boldsymbol{u}(\ell, k) \tag{4.1}$$

where $\boldsymbol{h}_d(\ell, k)$ and $\boldsymbol{h}_i(\ell, k)$ are the normalized ATFs relating the desired and interfering sources and the microphones, respectively, and $\boldsymbol{u}(\ell, k)$ is a spatially-white sensors noise. The

kth wavelength corresponding to the kth frequency index is $\lambda_k = \frac{2\pi c}{f_s}\frac{N_{\text{DFT}}}{k}$ where f_s is the sampling frequency and c is the sound velocity.

The desired and interfering ATFs are assumed time-invariant and are given by:

$$\boldsymbol{h}_d(\ell,k) = \frac{1}{\sqrt{M}}\left[\underbrace{1 \quad \cdots \quad 1}_{M}\right]^T \tag{4.2a}$$

$$\boldsymbol{h}_i(\ell,k) = \frac{1}{\sqrt{M}}\left[\, e^{-j\xi_i\frac{x_1}{\lambda_k}} \quad \cdots \quad e^{-j\xi_i\frac{x_M}{\lambda_k}} \,\right]^T \tag{4.2b}$$

where $\xi_i = 2\pi\sin(\theta_i)$. Note that by this notation the phase of both the desired source and of the interfering source is assumed to be 0 at the origin ($x = 0$). Note also that unlike the common notation, the ATFs in our work are normalized. For brevity, the frequency index is omitted and only the kth frequency index is considered. The same analysis is applicable to each frequency bin. Define the total interference by $\boldsymbol{v}(\ell) = \sqrt{M}\boldsymbol{h}_i s_i(\ell) + \boldsymbol{u}(\ell)$. The second moments of the received signals are denoted by:

$$\boldsymbol{\Phi}_{zz} = M\sigma_d^2\boldsymbol{h}_d\boldsymbol{h}_d^H + \boldsymbol{\Phi}_{vv} \tag{4.3}$$

where

$$\boldsymbol{\Phi}_{vv} = M\sigma_i^2\boldsymbol{h}_i\boldsymbol{h}_i^H + \sigma_u^2\boldsymbol{I} \tag{4.4}$$

is the covariance matrix of the total interference, $\sigma_d^2, \sigma_i^2, \sigma_u^2$ are the spectra in the kth frequency bin of the desired source, the interfering source and the microphone noise, respectively, and $\boldsymbol{I}$ is an $M \times M$ identity matrix. The goal of the MWF is to estimate a delayed version of the desired signal in the MMSE sense. Here, the desired signal is defined by the desired source component at the first microphone:

$$d(\ell) = \sqrt{M}h_{d,1}s_d(\ell). \tag{4.5}$$

The output of the beamformer $\boldsymbol{w}$ is denoted by $y_o(\ell) = \boldsymbol{w}^H\boldsymbol{z}(\ell)$.

4.1.2 MSE analysis given the microphone locations

As the microphone locations are random, the corresponding MWF and its corresponding MMSE are also RVs. Their statistics is analyzed in the following sections. In this section the

MSE of the MWF is analyzed for a given set of microphone locations. The MWF is given by:

$$\boldsymbol{w}_{\mathrm{MWF}} = \boldsymbol{\Phi}_{zz}^{-1}\boldsymbol{\phi}_{zd} \tag{4.6}$$

where

$$\boldsymbol{\phi}_{zd} = \mathrm{E}\left[\boldsymbol{z}(\ell)d^{*}(\ell)\right] \tag{4.7}$$

is the cross-correlation vector between the received signals and the desired signal. Substituting (4.1) in (4.7) gives

$$\boldsymbol{\phi}_{zd} = M\sigma_d^2 h_{d,1}\boldsymbol{h}_d. \tag{4.8}$$

For further simplification of (4.6), the Woodbury identity is applied to $\boldsymbol{\Phi}_{zz}$ in (4.3):

$$\begin{aligned}\boldsymbol{\Phi}_{zz}^{-1} =& \boldsymbol{\Phi}_{vv}^{-1} - \boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{h}_d\left(\left(M\sigma_d^2\right)^{-1} + \boldsymbol{h}_d^H\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{h}_d\right)^{-1}\boldsymbol{h}_d^H\boldsymbol{\Phi}_{vv}^{-1} \\ =& \left(\boldsymbol{I} - \left(\left(M\sigma_d^2\right)^{-1} + \alpha\right)^{-1}\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{h}_d\boldsymbol{h}_d^H\right)\boldsymbol{\Phi}_{vv}^{-1}\end{aligned} \tag{4.9}$$

where

$$\alpha = \boldsymbol{h}_d^H\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{h}_d. \tag{4.10}$$

Substituting (4.9) and (4.8) in (4.6) yields the expression

$$\boldsymbol{w}_{\mathrm{MWF}} = \frac{h_{d,1}}{\left(M\sigma_d^2\right)^{-1} + \alpha}\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{h}_d. \tag{4.11}$$

The corresponding MMSE is given by:

$$J_{\mathrm{MWF}} = \mathrm{E}\left[\left|d(\ell) - \boldsymbol{w}_{\mathrm{MWF}}{}^H\boldsymbol{z}(\ell)\right|^2\right] = \sigma_d^2 - \boldsymbol{w}_{\mathrm{MWF}}{}^H\boldsymbol{\Phi}_{zz}\boldsymbol{w}_{\mathrm{MWF}}. \tag{4.12}$$

Substituting (4.11) in (4.12) gives

$$J_{\mathrm{MWF}} = \sigma_d^2\frac{\left(M\sigma_d^2\right)^{-1}}{\left(M\sigma_d^2\right)^{-1} + \alpha}. \tag{4.13}$$

Denote the inner product of the desired and interfering ATFs by:

$$\rho = \boldsymbol{h}_d^H \boldsymbol{h}_i. \tag{4.14}$$

Applying the Woodbury identity to $\boldsymbol{\Phi}_{vv}$ in (4.4) and substituting the result in (4.10) yields

$$\alpha = \frac{1}{\sigma_u^2}\left(1 - \frac{\sigma_i^2}{\sigma_i^2 + \sigma_u^2/M}|\rho|^2\right). \tag{4.15}$$

Finally, denoting the following spectra ratios

$$\gamma_d = \frac{\sigma_d^2}{\sigma_d^2 + \sigma_u^2/M} \tag{4.16a}$$

$$\gamma_i = \frac{\sigma_i^2}{\sigma_i^2 + \sigma_u^2/M} \tag{4.16b}$$

and substituting (4.15) in (4.13) gives

$$J_{\mathrm{MWF}}\left(|\rho|^2\right) = \frac{\sigma_u^2}{M}\gamma_d\left(1 - \gamma_d\gamma_i|\rho|^2\right)^{-1}. \tag{4.17}$$

The last expression for the MMSE depends only on the sources spectra, the number of microphones, and ρ. The residual error power, J_{MWF}, is a monotonically increasing function of $|\rho|^2$. Considering that $0 \leq |\rho|^2 \leq 1$ and the monotonic behavior of J_{MWF}, it is bounded by $\sigma_d^2\frac{\sigma_u^2/M}{\sigma_d^2+\sigma_u^2/M} \leq J_{\mathrm{MWF}} \leq \sigma_d^2\frac{\sigma_i^2+\sigma_u^2/M}{\sigma_d^2+\sigma_i^2+\sigma_u^2/M}$. The lower bound corresponds to the case where the sources' ATFs are orthogonal. Its corresponding MWF is the average of M identical single channel Wiener filters calculated at the absence of the interference. The upper bound corresponds to the case where the sources' ATFs coincide. In this case the residual error power is maximized since it is impossible to spatially separate the desired and interfering sources. Its corresponding MWF is the average of M identical single channel Wiener filters assuming that the interference is present.

In Sec. 4.1.3 the microphone locations are assumed random and the statistics of ρ is analyzed. The statistics of ρ will then be used for analyzing the statistics of J_{MWF} based on (4.17) .

4.1.3 The statistics of ρ

A summary of the statistics of ρ which was derived by Lo in [6] follows. ρ is a complex RV, $\rho = \rho_r + \mathrm{i}\rho_i$, with real and imaginary components denoted by ρ_r and ρ_i. Using (4.14) and (4.2a,4.2b) it can be verified that

$$\rho_r = \frac{1}{M}\sum_{m=1}^{M} \cos\left(\xi_i \frac{x_m}{\lambda}\right) \tag{4.18a}$$

$$\rho_i = \frac{1}{M}\sum_{m=1}^{M} \sin\left(\xi_i \frac{x_m}{\lambda}\right) \tag{4.18b}$$

where $\mathrm{i} = \sqrt{-1}$. Now since $\{x_m\}_{m=1}^{M}$ are i.i.d. RVs, it can be shown that the first-order and second-order moments of ρ_r and ρ_i are given by:

$$\mu_{\rho,r} = \mathrm{E}\left[\rho_r\right] = \phi_x(\frac{\xi_i}{\lambda}) \tag{4.19a}$$

$$\sigma^2_{\rho,r} = \mathrm{E}\left[(\rho_r - \mu_{\rho,r})^2\right] = \frac{1}{2M}\left(1 + \phi_x(2\frac{\xi_i}{\lambda})\right) - \phi_x^2(\frac{\xi_i}{\lambda}) \tag{4.19b}$$

$$\mu_{\rho,i} = \mathrm{E}\left[\rho_i\right] = 0 \tag{4.19c}$$

$$\sigma^2_{\rho,i} = \mathrm{E}\left[(\rho_i - \mu_{\rho,i})^2\right] = \frac{1}{2M}\left(1 - \phi_x(2\frac{\xi_i}{\lambda})\right) - \phi_x^2(\frac{\xi_i}{\lambda}) \tag{4.19d}$$

where $\phi_x(t) = \mathrm{sinc}(t\Delta)$ denotes the characteristic function of the RV x, in the Uniform distribution case. The summands of the summation in ρ_r, ρ_i are i.i.d. RVs. Therefore, according to the central limit theorem (CLT) they converge to a Gaussian RV for $M \gg 1$. Assuming that

$$\frac{\xi_i \Delta}{\lambda} \gg 1 \tag{4.20}$$

the following approximation holds:

$$\rho_r \sim \mathcal{N}(0, \frac{1}{2M}) \tag{4.21a}$$

$$\rho_i \sim \mathcal{N}(0, \frac{1}{2M}). \tag{4.21b}$$

Note that this approximation is not valid for $\theta_i \to 0$. Since ρ_r and ρ_i are uncorrelated, i.e. $\mathrm{E}\left[\rho_r \rho_i\right] = 0$, $2M|\rho|^2 = 2M{\rho_r}^2 + 2M{\rho_i}^2$ is approximated by a χ^2 RV with 2 degrees of freedom,

which is an Exponential RV with a parameter 1/2, i.e.

$$2M|\rho|^2 \sim \exp(1/2). \tag{4.22}$$

Note that when assumption (4.20) holds, ρ tends to its lower bound 0 as the number of microphones M increases, and according to (4.17) the corresponding J_{MWF} tends to its lower bound as well.

4.1.4 Statistics of the MSE

Using the Exponential distribution of $|\rho|^2$, the expression for J_{MWF} in (4.17) and its monotonic behaviour, the cumulative distribution function (c.d.f.) of J_{MWF} is given by:

$$\begin{aligned}
&\Pr\left(J_{\mathrm{MWF}}(|\rho|^2) \leq J_0\right) = \Pr\left(|\rho|^2 \leq J_{\mathrm{MWF}}^{-1}\left(J_0\right)\right) = \\
&\Pr\left(|\rho|^2 \leq {\gamma_d}^{-1}{\gamma_i}^{-1}\left(1 - \frac{\sigma_u^2/M}{J_0}\gamma_d\right)\right) = \\
&1 - \exp\left(-M{\gamma_d}^{-1}{\gamma_i}^{-1}\left(1 - \frac{\sigma_u^2/M}{J_0}\gamma_d\right)\right).
\end{aligned} \tag{4.23}$$

Equation (4.23) denotes a reliability measure of J_{MWF}. It equals the probability that the MMSE will not exceed a desired level, J_0.

4.1.5 Model verification

We turn now to the verification of the derived models. We verify the Normal model of ρ in (4.21a,4.21b), and the Exponential models of $|\rho|^2$ and the reliability function in (4.22) and (4.23), respectively. For each scenario a Monte-Carlo simulation consisting of 1000 arrays of $M = 21$ microphones (unless stated otherwise) were randomized with a Uniform distribution on a linear aperture of $\Delta = 10$ length units. A desired source and an interfering source arriving from the far-field were simulated. The angle of arrival (AOA) of the desired source was set to 0^o, and the AOA of the interference was set to $\theta_i = 5.5^o$ (unless stated otherwise). A low level sensors noise was added to the received signals. The SIR was set to 0dB and the SNR was set to 30dB. The results were obtained for wavelengths in the range of $[0.1, 2]$ length units, and are shown for a a specific wavelength of $\lambda = 0.91$ length units (unless stated otherwise).

4.1.5.1 The Normal model of the components of ρ and the Exponential model of $|\rho|^2$

The Normal probability plots of ρ_r and the Exponential probability plots of $2M|\rho|^2$ were simulated for various combinations of λ and θ_i and the results are shown in Fig. 4.1(a) and Fig. 4.1(b), respectively. The blue color corresponds to $\theta_i = 1.5^o, \lambda = 0.41 \Rightarrow \xi_i\frac{\Delta}{\lambda} = 0.64$, the green color corresponds to $\theta_i = 9.5^o, \lambda = 2 \Rightarrow \xi_i\frac{\Delta}{\lambda} = 0.82$, and the red color corresponds to $\theta_i = 9.5^o, \lambda = 0.41 \Rightarrow \xi_i\frac{\Delta}{\lambda} = 4.02$. The various markers (cross, circle and plus) correspond to the measurements data and the dashed lines correspond to the best distribution fit. Markers that coincide with the dashed lines correspond to a good fit between the distribution model and the data.

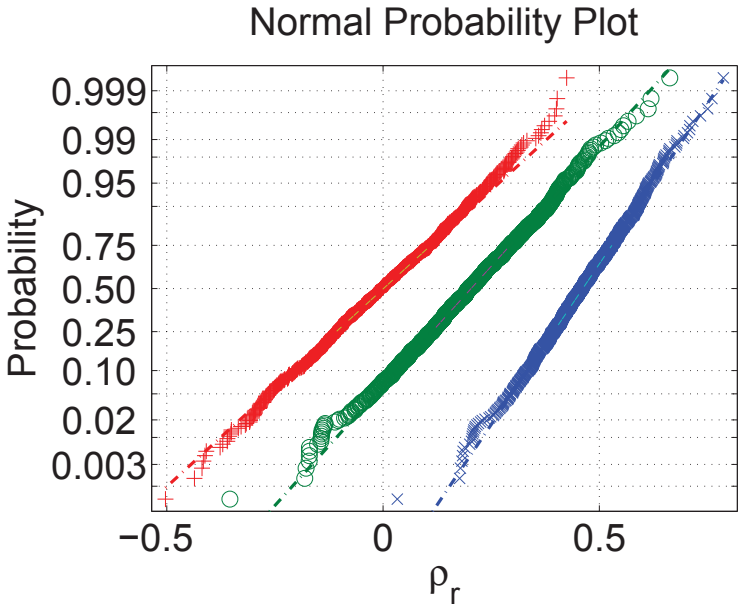
Since assumption (4.20) holds only for the parameters of the red curve, we expect it to coincide with the theoretical distribution models in (4.21a,4.21b) and (4.22). It is clear from Fig. 4.1(a) that a Normal distribution model fits the blue, green and red curves. However, only the red curve matches the zero mean model of ρ in (4.21a,4.21b). From Fig. 4.1(b) it is obvious that only the red curve matches the Exponential distribution. The estimated mean and variance of ρ_r are -3.1×10^{-3} and 2.34×10^{-2}. These values match the theoretical values of 0 and $2.38\times10^{-2} = \frac{1}{2\times21}$ in (4.21a,4.21b). The estimated parameter of the Exponential distribution of $2M|\rho|^2$ is 0.49. It matches its theoretical value of 0.5 in (4.22). The blue and the green curves are examples for cases where assumption (4.20) is invalid, rendering the Exponential distribution inappropriate for representing the data points.

4.1.5.2 The reliability of $J_{\mathbf{MWF}}$

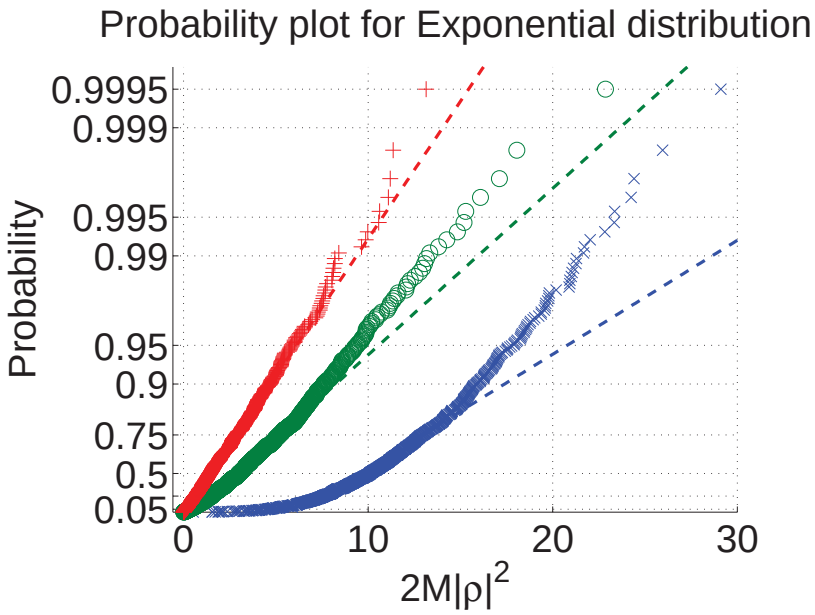
In order to verify the reliability of J_{MWF} the number of microphones was set in the range $M = 5, 6, \ldots, 30$. The analytical and empirical reliability functions for $M = 11$ microphones are depicted in Fig. 4.2(a). It is clear from this figure that the theoretical model in (4.23) matches the empirical data. The MMSE normalized by the signal power in this case varies in the range of -41dB to -36dB.

The theoretical value of the c.d.f. of the MMSE at point $\Pr\left(10\log\left(\frac{J_{\mathrm{MWF}}}{\sigma_d^2}\right) \leq -40\right)$, as well as its empirical value are depicted in Fig. 4.2(b). It is clear from this figure that the theoretical model matches the empirical results. It is evident that the c.d.f. has an approximate step function characteristics. Below a certain threshold (number of microphones) the c.d.f. tends to 0 and above the threshold it tends to 1 with an abrupt transition between the

two values. The threshold in Fig. 4.2(b) is approximately at $M = 11$ microphones. Using more than 11 microphones, it is almost guaranteed that the MWF error will be lower than -40dB.

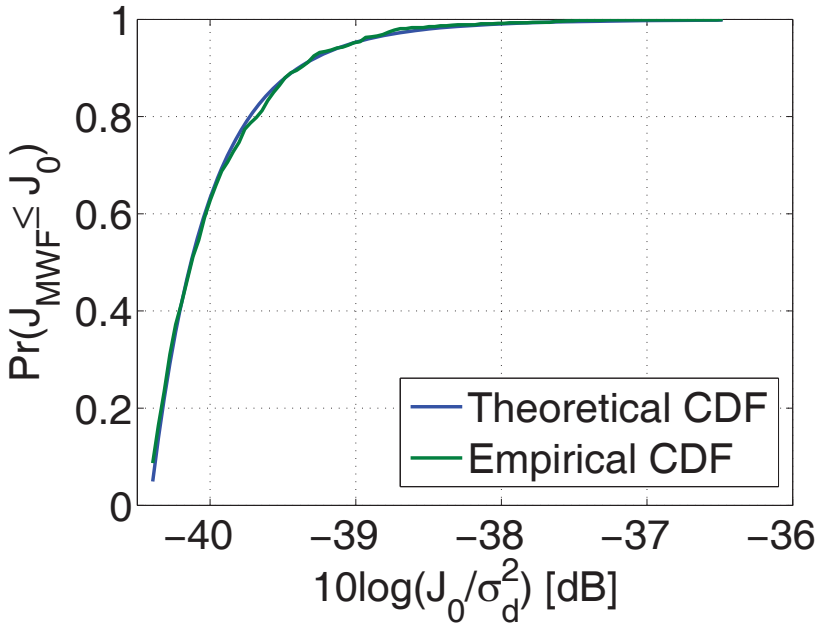


(a) Normal probability plots of ρ_r.

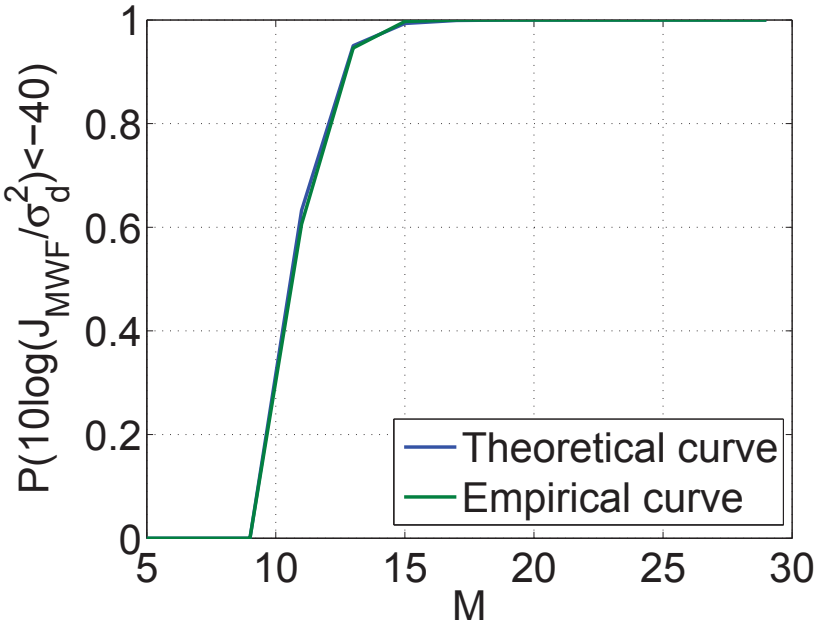


(b) Exponential probability plots of $2M|\rho|^2$.

Figure 4.1: Normal and Exponential probability plots of ρ_r and $|\rho|^2$, respectively, for various values of $\xi_i \frac{\Delta}{\lambda}$ (0.64 in blue, 0.82 in green and 4.02 in red).



(a) Theoretical (blue) and empirical (green) c.d.f. curves of J_{MWF} with an $M = 11$ microphones array.

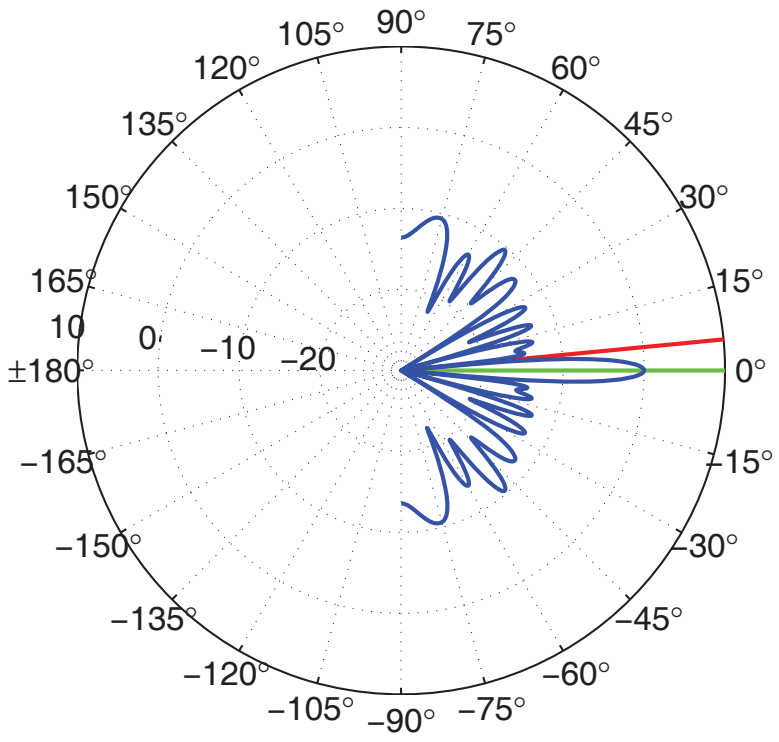


(b) Theoretical (blue) and empirical (green) values of $\Pr\left(10\log\left(\frac{J_{\text{MWF}}}{\sigma_d^2}\right) \leq -40\right)$ vs. M, the number of microphones.

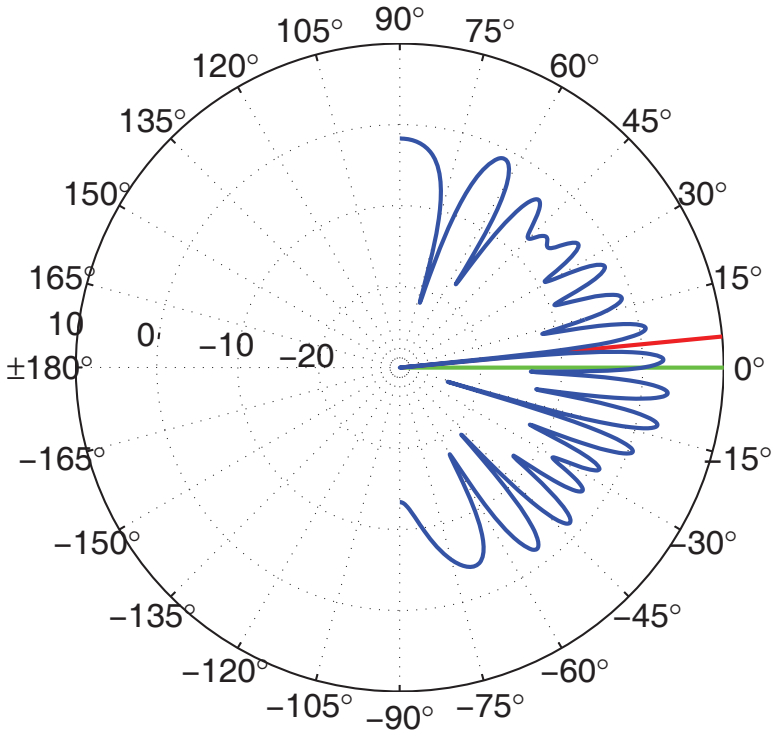
Figure 4.2: The c.d.f. of J_{MWF} as a function of J_0 and M, the number of microphones.

Examples for a beampattern corresponding to an array realization with $M = 9$ micro-

phones (below the threshold), and for a beampattern corresponding to an array realization with $M = 15$ microphones (above the threshold) are depicted in Fig. 4.3(a) and Fig. 4.3(b), respectively. The beampattern in Fig. 4.3(a) has low sidelobes while the beampattern in Fig. 4.3(b) seems to suffer from spatial aliasing and as a results, exhibits high sidelobes.



(a) An example beampattern with $M = 15$ microphones.



(b) An example beampattern with $M = 9$ microphones.

Figure 4.3: Examples for beampatterns for two array realizations. The AOAs of the desired and interfering sources are depicted in green and red, respectively.

4.1.6 Conclusions

In many applications, e.g. border security, exact design of the array is impractical. The performance of these arrays, characterized by a large number of microphones deployed in vast areas, cannot be analyzed by traditional deterministic methods. In the current contribution we have presented a statistical model for the performance of a randomly located linear microphone array. Specifically, we analyzed the MMSE measure of the MWF, commonly used in speech applications. The case of one desired source and one interfering source arriving from the far-field was treated. The theoretical models that have been developed were verified by simulations. The proposed model can be used for determining the number of microphones required for obtaining a predefined residual error level.

Special considerations need to be made when processing speech signals. Throughout this work we assumed that the second-order statistics of the various sources is available. This is hardly ever the case in actual applications. A practical design of the MWF for speech processing is given by Doclo and Moonen [29]. The derivation of the reliability measure assumes that the sources are stationary with known spectra. The reliability measure can be extended to the case of speech signals in several ways. The stationary spectra can be replaced by instantaneous spectra estimates, or time averaged spectra. Alternatively, the required number of microphones can be determined by worst-case considerations.

4.2 Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure

In this section, we consider an array of randomly located microphones with a uniform distribution in a reverberant enclosure. A single desired speaker is assumed. Utilizing the statistical model of the RIR, we derive a statistical model for the SIR and white noise gain and introduce their reliability functions for the SDW-MWF in several noise scenarios. Specifically, we consider the case of $P < M$ coherent noise sources, where M is the number of microphones, and the case of a diffuse sound field, an infinite number of uncorrelated sources arriving from all directions. The derived statistical model, and the reliability functions can be used to determine the number of microphones needed to assure a desired performance level (with a controlled level of uncertainty). Note, that we extend the contribution of the previous Sec. 4.1 in two aspects: first, a reverberant environment is treated instead of a far-field direct

arrival only model; second, we analyze the performance for both coherent and diffuse sound fields.

The structure of the section is as follows. The problem is formulated in Sec. 4.2.1. We briefly present the SDW-MWF, and derive expressions for SIR and white noise gain in Sec. 4.2.2. In Sec. 4.2.3 we derive a statistical model for the ATF, and experimentally verify its validity. Sec. 4.2.4 is dedicated to deriving a statistical model for the BF criteria, namely SIR and white noise gain, in various noise fields: 1) few coherent noise sources $P < M$; 2) diffuse sound field. In Sec. 4.2.5 we compare the proposed model and the empirical results obtained by an extensive simulative study. The work is summarized and conclusions are drawn Sec. 4.2.6.

4.2.1 Problem formulation

In the current contribution we analyze the performance of the SDW-MWF, aiming to enhance a desired source contaminated by interference sources. We consider the case of randomly located microphones with a uniform distribution in a reverberant enclosure, and derive a statistical model for the performance.

Consider a $D_x \times D_y \times D_z$ dimensional reverberant room, in which M microphones are randomly located with a uniform distribution. The microphone locations are given in a Cartesian coordinate system, with the origin at the center of the room:

$$\boldsymbol{r}^m \triangleq \left[\begin{array}{ccc} r_x^m & r_y^m & r_z^m \end{array} \right]^T \tag{4.24}$$

for $m = 1, \ldots, M$, where $(\bullet)^T$ denotes the transpose operator. Throughout this contribution, the STFT domain is considered, where ℓ and k denote the frame and frequency indices. Let $s_d(\ell, k)$ be a desired speaker positioned at $\boldsymbol{r}_d$ in the enclosure. The signals received by the microphones are given by:

$$\boldsymbol{z}(\ell, k) \triangleq \boldsymbol{h}_d(\ell, k)\, s_d(\ell, k) + \boldsymbol{v}(\ell, k) \tag{4.25}$$

where $\boldsymbol{h}_d(\ell, k) \triangleq \left[\begin{array}{ccc} h_d^1(\ell, k) & \cdots & h_d^M(\ell, k) \end{array} \right]^T$ denotes the ATF relating the desired speech signal and the microphones, and $\boldsymbol{v}(\ell, k)$ is a vector comprised of the interference signals. At this point, a general noise field is assumed. In the next sections several specific cases are addressed, explicitly, $P < M$ coherent interference sources as well as a diffuse sound

field are considered. Denote the received signals, and the interference covariance matrices, respectively, as:

$$\begin{aligned}\boldsymbol{\Phi}_{zz}(\ell,k) \triangleq & \mathrm{E}\left\{\boldsymbol{z}(\ell,k)\,\boldsymbol{z}^{H}(\ell,k)\right\} \\ = & \delta_d(\ell,k)\,\boldsymbol{h}_d(\ell,k)\,\boldsymbol{h}_d^{H}(\ell,k)+\boldsymbol{\Phi}_{vv}(\ell,k) \qquad (4.26a)\\ \boldsymbol{\Phi}_{vv}(\ell,k) \triangleq & \mathrm{E}\left\{\boldsymbol{v}(\ell,k)\,\boldsymbol{v}^{H}(\ell,k)\right\} \qquad (4.26b)\end{aligned}$$

where $\mathrm{E}\{\bullet\}$ denotes the expectation operator, $(\bullet)^H$ denotes the transpose-conjugate operator, and $\delta_d(\ell,k) \triangleq \mathrm{E}\{|s_d(\ell,k)|^2\}$ denotes the PSD of the desired signal. We assume that the desired speech signal and the interference signals are statistically independent. For brevity, hereafter the frame index ℓ is omitted from the covariance matrices and the ATFs of the desired and interfering sources. The frequency index $k = 1,\ldots,K$ where K is the window length, is also omitted, and the subsequent derivations should be understood as frequency dependent.

In the following section, the SDW-MWF is briefly presented, and its performance criteria are defined.

4.2.2 The SDW-MWF

The MSE between the output of a BF, $\boldsymbol{w}'$, and the desired signal is $\mathrm{E}\left\{|s_d(\ell) - (\boldsymbol{w}')^H\,\boldsymbol{z}(\ell)|^2\right\}$. The SDW-MWF BF is designed to minimize a weighted version of the MSE and its goal is to enhance the desired signal $s_d(\ell)$. It is defined as the solution of the following minimization problem:

$$\begin{aligned}\boldsymbol{w} \triangleq & \underset{\boldsymbol{w}'}{\operatorname{argmin}}\,|1-\left((\boldsymbol{w}')^H\,\boldsymbol{h}_d\right)|^2\delta_d+\mu\,(\boldsymbol{w}')^H\,\boldsymbol{\Phi}_{vv}\boldsymbol{w}' \\ = & \frac{\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{h}_d}{\boldsymbol{h}_d^{H}\boldsymbol{\Phi}_{vv}^{-1}\boldsymbol{h}_d+\frac{\mu}{\delta_d}} \qquad (4.27)\end{aligned}$$

where μ is a non-negative parameter which controls the tradeoff between the interference reduction and the desired signal distortion. For $\mu = 1$ the classical Wiener filter [25] (MMSE) is obtained. At the limit $\mu \to 0$ the MVDR-BF is reached, and no distortion is introduced to the desired signal.

Next, we define two criteria of BFs. The SIR at the output of a BF, $\boldsymbol{w}$, is denoted κ and is defined as the ratio of the powers of the desired signal and the interference signals at the

beamformer output, i.e.:

$$\kappa \triangleq \frac{\delta_d |\boldsymbol{w}^H \boldsymbol{h}_d|^2}{\boldsymbol{w}^H \boldsymbol{\Phi}_{vv} \boldsymbol{w}}. \tag{4.28}$$

The white noise gain [27] is denoted ξ and is defined as the SNR gain of the BF for a spatially white noise. It equals:

$$\xi \triangleq \frac{|\boldsymbol{w}^H \boldsymbol{h}_d|^2}{\|\boldsymbol{w}\|^2}. \tag{4.29}$$

By substituting (4.27) in (4.28) and (4.29), we obtain expressions for the performance criteria of the SDW-MWF:

$$\kappa = \delta_d \boldsymbol{h}_d^H \boldsymbol{\Phi}_{vv}^{-1} \boldsymbol{h}_d. \tag{4.30a}$$

$$\xi = \frac{\left(\boldsymbol{h}_d^H \boldsymbol{\Phi}_{vv}^{-1} \boldsymbol{h}_d\right)^2}{\boldsymbol{h}_d^H \boldsymbol{\Phi}_{vv}^{-2} \boldsymbol{h}_d}. \tag{4.30b}$$

Note, that both expressions do not depend on μ. This can be attributed to the fact that SDW-MWF equals the MVDR-BF followed by a single channel SDW-MWF [110],[31] with the parameter μ. Hence, the SIR and white noise gain at the output of the SDW-MWF equal to their respective quantities at the output of an MVDR-BF (locally, per frequency bin).

In the next section, a statistical model for the ATF is presented. From this model we will derive the statistics of the SIR and the white noise gain criteria for various noise fields.

4.2.3 ATF statistics

In the following sections approximations for the first and second moments of the ATF are derived. We will show that, under several assumptions, the ATFs relating a source with the microphone array can be modeled as i.i.d. complex Gaussian RVs with zero mean and a variance which depends on the properties of the enclosure. Furthermore, ATFs of different sources are shown to be uncorrelated.

4.2.3.1 Single ATF statistics

The performance of statistical BFs relies greatly on the acoustic propagation model. Considering acoustic signals propagation, Schroeder, in the 1950s, proposed a stochastic model for the RIR and the respective frequency correlation between microphones. This work was

later translated to English in [111] and further developed by Polack [112]. In [113], Schroeder investigated the frequency correlation between frequency responses in a room. For a survey on the topic please refer to Kuttroff [114] and to Jot et al. [115].

Let h be an ATF relating a coherent point source signal, located at $\boldsymbol{r}$, and the mth microphone, located at $\boldsymbol{r}^m$. The ATF is comprised of two components: the direct arrival ATF and the reverberant component ATF. Since the microphone location, $\boldsymbol{r}^m$, is random, the ATF is also a complex RV. Explicitly:

$$h \triangleq \bar{h} + \hat{h} \tag{4.31}$$

where $\bar{h}$ and $\hat{h}$ denote direct arrival and reverberant components, respectively. We assume that the direct arrival and the reverberant ATFs are uncorrelated. Define the room volume and surface area as:

$$V \triangleq D_x \times D_y \times D_z \tag{4.32a}$$

$$A \triangleq 2\left(D_x \times D_y + D_x \times D_z + D_y \times D_z\right) \tag{4.32b}$$

and denote the reverberation time as T_{60}. Adopting the ATF model of Schroeder [111],[114], $\hat{h}$ is modeled as a complex Gaussian RV:

$$\hat{h} \sim \mathcal{CN}\left(0, \hat{\alpha}\right) \tag{4.33}$$

where

$$\hat{\alpha} \triangleq \frac{1-\varepsilon}{\pi \varepsilon A} \tag{4.34}$$

and

$$\varepsilon \triangleq \frac{0.161V}{AT_{60}} \tag{4.35}$$

is the exponential decay rate of the RIR tail. The latter model is valid under the following assumptions: first, the signal wavelength is much smaller than the room dimensions; second, microphones and sources are at least half a wavelength away from the walls; third, the signal

frequency is above the Schroeder frequency, defined as:

$$f_{\text{Schroeder}} \triangleq 2000\sqrt{\frac{T_{60}}{V}}. \tag{4.36}$$

In typical acoustic scenarios $f_{\text{Schroeder}}$ is in the order of few hundred Hz.

The direct arrival ATF is given in a polar representation by

$$\bar{h} \triangleq \bar{a}\exp\left(\mathrm{j}\bar{\phi}\right) \tag{4.37}$$

where $\bar{a}$ and $\exp\left(\mathrm{j}\bar{\phi}\right)$ are the amplitude and phase responses, respectively. Assuming spherical wave propagation:

$$\bar{a} = \begin{cases} 1 & ;|\boldsymbol{r}| \leq \frac{1}{4\pi} \\ \frac{1}{4\pi\|\boldsymbol{r}-\boldsymbol{r}^m\|} & ;\frac{1}{4\pi} < |\boldsymbol{r}| \end{cases} \tag{4.38a}$$

$$\bar{\phi} = \frac{2\pi\|\boldsymbol{r}-\boldsymbol{r}^m\|}{\lambda_k} \tag{4.38b}$$

where $\lambda_k = \frac{cK}{kf_s}$ is the wavelength corresponding to the kth frequency bin, K is the STFT window length, f_s is the sampling frequency rate and c is the sound velocity. Furthermore, since (4.38a) is not physically meaningful for $\|\boldsymbol{r}-\boldsymbol{r}^m\| \to 0$, we limit the amplitude response to $\bar{a} = 1$ for $\|\boldsymbol{r}-\boldsymbol{r}^m\| < \frac{1}{4\pi}$. Without loss of generality, consider that the source is located at the origin $\boldsymbol{r} = \boldsymbol{0}$, and that a sphere with radius $\bar{r}$, centered at the origin, is within the room volume. The sphere radius is chosen such that $\bar{r} \gg \lambda_k$. Since multiple 2π phase cycles are repeated while propagating in the sphere, the amplitude and phase responses, $\bar{a}$ and $\exp\left(\mathrm{j}\bar{\phi}\right)$ can be approximated as uncorrelated inside the sphere. We verify this approximation in Sec. 4.2.3.3. Note, that the direct arrival component is a stochastic variable, since it is a function of the microphone location which is random, and that the reverberant component is stochastic under Schroeder's model. The subsequent expectation operations should be interpreted accordingly.

Moreover, the mean phase response is approximately

$$\mathrm{E}\left\{\exp\left(\mathrm{j}\bar{\phi}\right)\right\} \approx 0 \tag{4.39}$$

and the mean direct arrival ATF is:

$$\begin{aligned}\mathrm{E}\left\{\bar{h}\right\} =&\mathrm{E}\left\{\bar{a}\right\}\mathrm{E}\left\{\exp\left(\mathrm{j}\bar{\phi}\right)\right\}\\ =&0.\end{aligned} \tag{4.40}$$

These results are also verified in Sec. 4.2.3.3.

Considering a microphone which location is uniformly distributed in a sphere with radius $\bar{r}$ and using (4.38a), the variance of the direct arrival equals the solution of the following integral in spherical coordinates:

$$\begin{aligned}\bar{\alpha} =&\mathrm{E}\left\{\bar{a}^2\left|\|\boldsymbol{r}^m\|<\bar{r}\right.\right\}\\ =&\int_{\phi=0}^{2\pi}\int_{\theta=0}^{\pi}\int_{r=0}^{\bar{r}}\frac{3}{4\pi\bar{r}^3}\bar{a}^2r^2\sin\left(\theta\right)\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi\\ =&\frac{6\pi\bar{r}-1}{32\pi^3\bar{r}^3}.\end{aligned} \tag{4.41}$$

Combining (4.33) and (4.40), the mean ATF equals:

$$\mathrm{E}\left\{h\right\}=0. \tag{4.42}$$

Denote the variance of the ATF:

$$\alpha\triangleq\mathrm{E}\left\{|h|^2\right\}. \tag{4.43}$$

Using the law of total probability, (4.43) can be written as:

$$\begin{aligned}\alpha =&\Pr\left(\|\boldsymbol{r}^m\|<\bar{r}\right)\mathrm{E}\left\{|h|^2\left|\|\boldsymbol{r}^m\|<\bar{r}\right.\right\}\\ &+\Pr\left(\|\boldsymbol{r}^m\|\geq\bar{r}\right)\mathrm{E}\left\{|h|^2\left|\bar{r}\leq\|\boldsymbol{r}^m\|\right.\right\}.\end{aligned} \tag{4.44}$$

Denote the critical distance, the distance from the source at which the powers of the direct arrival and the reverberant components are equal, as r_c. Kuttruff [114] derived an expression for the critical distance:

$$r_c\triangleq\sqrt{\frac{V}{100\pi T_{60}}}. \tag{4.45}$$

We assume that $\bar{r}\gg r_c$, i.e. the sphere radius $\bar{r}$ is much larger than the critical distance.

Hence $|\bar{h}| \ll |\hat{h}|$ (the direct arrival ATF is negligible compared to the reverberant ATF component) for $\|\boldsymbol{r}^m\| > \bar{r}$, and (4.44) is approximately:

$$\begin{aligned}\alpha \approx &\Pr\left(\|\boldsymbol{r}^m\| < \bar{r}\right) \mathrm{E}\left\{|\bar{h}|^2 + |\hat{h}|^2 \,|\|\boldsymbol{r}^m\| < \bar{r}\right\} \\ &+ \Pr\left(\|\boldsymbol{r}^m\| \geq \bar{r}\right) \mathrm{E}\left\{|\hat{h}|^2 \,|\bar{r} \leq \|\boldsymbol{r}^m\|\right\} \\ = &\Pr\left(\|\boldsymbol{r}^m\| < \bar{r}\right) \mathrm{E}\left\{|\bar{h}|^2 \,|\|\boldsymbol{r}^m\| < \bar{r}\right\} + \mathrm{E}\left\{|\hat{h}|^2\right\}. \end{aligned} \tag{4.46}$$

Note, that we utilized the fact that $\bar{h}$ and $\hat{h}$ are uncorrelated. Substituting (4.41) and (4.33) in (4.46) and noting that $\Pr\left(\|\boldsymbol{r}^m\| < \bar{r}\right) = \frac{4\pi\bar{r}^3}{3V}$ yields:

$$\alpha = \frac{4\pi\bar{r}^3}{3V}\bar{\alpha} + \hat{\alpha}. \tag{4.47}$$

We return to the original coordinate system, centered in the room, and consider the statistics of an ATF vector $\boldsymbol{h}$. Since the microphone locations are i.i.d., and since h^m, the mth element in $\boldsymbol{h}$, depends on the location of the mth microphone, $\boldsymbol{r}^m$, we conclude that $h^m;\ m = 1, \ldots, M$ are i.i.d..

4.2.3.2 Cross -covariance of ATFs

In this section we model the cross-covariance of the ATFs relating two sources located at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ with a microphone randomly located at $\boldsymbol{r}^m$. The covariance is comprised of the sum of the direct arrival and reverberant component covariances:

$$\mathrm{E}\left\{h_1 h_2^*\right\} = \mathrm{E}\left\{\bar{h}_1 \bar{h}_2^*\right\} + \mathrm{E}\left\{\hat{h}_1 \hat{h}_2^*\right\} \tag{4.48}$$

where h_i, $\bar{h}_i$ and $\hat{h}_i$ are, respectively, the total ATF, the direct arrival ATF and the reverberant ATF for sources $i = 1, 2$. Similarly to (4.37), the amplitude and phase components of the ith source direct arrival ATF are defined as $\bar{h}_i = \bar{a}_i \exp\left(\mathrm{j}\bar{\phi}_i\right)$.

First, let us examine the covariance of the reverberant ATFs. Schroeder models the reverberant ATF as the sum of multiple independent reflections arriving from all directions. Hence, their coherence, defined as $\frac{\mathrm{E}\left\{\hat{h}_1 \hat{h}_2^*\right\}}{\hat{\alpha}}$, is equivalent to the coherence between two microphones located at $\boldsymbol{r}_1$ and $\boldsymbol{r}_2$ in a diffuse sound field (comprised of multiple uncorrelated

sources radiating from a surrounding sphere) [116]. Explicitly, the covariance equals:

$$\mathrm{E}\left\{\hat{h}_1\hat{h}_2^*\right\} = \hat{\alpha}\,\mathrm{sinc}\left(\frac{2\pi\|\boldsymbol{r}_1-\boldsymbol{r}_2\|}{\lambda_k}\right) \tag{4.49}$$

where $\mathrm{sinc}(x)=\frac{\sin(x)}{x}$. Assuming that the distance between the sources is much larger than the wavelength, i.e. $\|\boldsymbol{r}_1-\boldsymbol{r}_2\| \gg \lambda_k$, the covariance between the reverberant component is approximately:

$$\mathrm{E}\left\{\hat{h}_1\hat{h}_2^*\right\} \approx 0. \tag{4.50}$$

Consider the expectation of the inner product of direct arrival ATFs in (4.48), i.e. $\mathrm{E}\left\{\bar{h}_1\bar{h}_2^*\right\} = \mathrm{E}\left\{\bar{a}_1\bar{a}_2\exp\left(\mathrm{j}\left(\bar{\phi}_1-\bar{\phi}_2\right)\right)\right\}$. Again, since $\|\boldsymbol{r}_1-\boldsymbol{r}_2\| \gg \lambda_k$ and by applying same considerations as leading to (4.40), we approximate that the phase and amplitude in the last expression are uncorrelated, and that $\mathrm{E}\left\{\exp\left(\mathrm{j}\left(\bar{\phi}_1-\bar{\phi}_2\right)\right)\right\} \approx 0$. Hence:

$$\mathrm{E}\left\{\bar{h}_1\bar{h}_2^*\right\} \approx 0. \tag{4.51}$$

Finally, substituting (4.50) and (4.51) in (4.48), we conclude that the ATFs are uncorrelated:

$$\mathrm{E}\left\{h_1h_2^*\right\} = 0. \tag{4.52}$$

4.2.3.3 Model verification

In order to verify the proposed simplified model we have conducted several different Monte-Carlo experiments. First, the model of the ATF statistics is examined for various reverberation times and enclosure dimensions. The theoretical model was calculated with the parameter $\bar{r}=2r_c$. In the first experiment, the location of a single microphone was uniformly randomized in a $4\mathrm{m}\times 4\mathrm{m}\times 3\mathrm{m}$ room with a reverberation time in the range $[0.2\mathrm{s}, 0.3\mathrm{s}, \ldots, 0.8\mathrm{s}]$. In the second experiment, we examined the relation between the ATF statistics and the room dimensions. We set the reverberation time to 0.3s and examined different room sizes: $(2+0.5i)\mathrm{m}\times(2+0.5i)\mathrm{m}\times(2.2+0.1i)\mathrm{m}$ for $i=0,\ldots,8$. For each room configuration in both experiments, the locations of a single source and a microphone were uniformly randomized in the room. The locations of the single source were randomized 4 times, and for each case 100 locations of the microphone were randomized. For each case, the direct arrival and the tail of the ATF as well as the complete ATF were generated using [108]. The normalized

error of the empirical mean of the ATF, is defined as $\frac{|\langle h\rangle - \mathrm{E}\{h\}|^2}{\mathrm{E}\{|h|^2\}} = \frac{|\langle h\rangle|^2}{\mathrm{E}\{|h|^2\}}$, where $\langle \bullet \rangle$ denotes the empirical average. As we expect that $\mathrm{E}\{h\} \approx 0$, as in Eq. (4.42), the ratio $\frac{|\langle h\rangle|^2}{\mathrm{E}\{|h|^2\}}$ is considered to verify this approximation in both experiments. For all tested reverberation times and room dimensions the normalized error, averaged over all considered scenarios, is -20dB, and clearly the approximation $\mathrm{E}\{h\} \approx 0$ holds. The theoretical and empirical (denoted

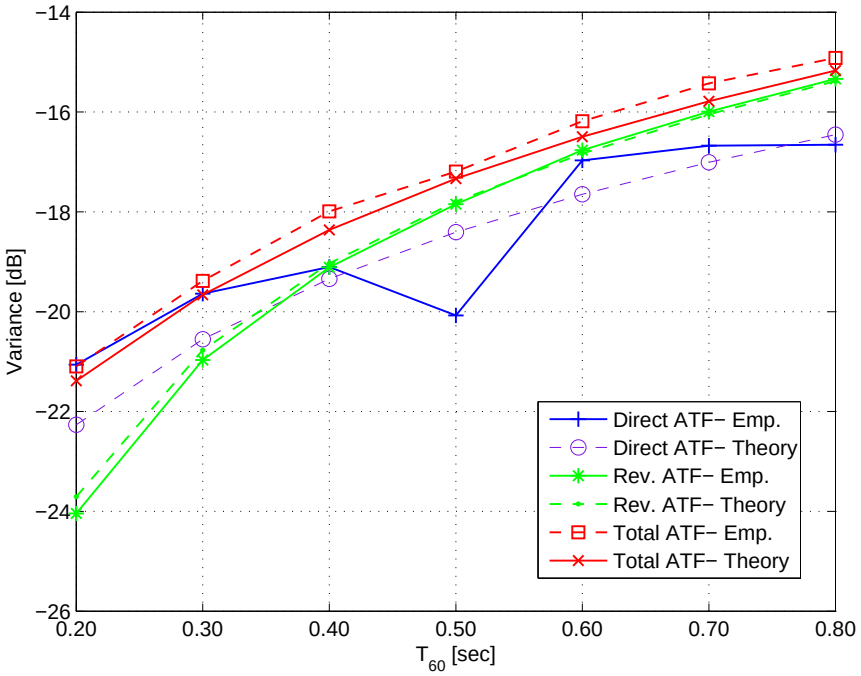


Figure 4.4: Empirical and theoretical ATF variances versus reverberation time.

emp.) variances of direct arrival, reverberant (denoted rev.) and total ATFs, i.e. $\bar{\alpha}$, $\hat{\alpha}$ and α, are depicted in Figs. 4.4,4.5 for both experiments, respectively. The empirical variances were averaged over the frequency range of [300Hz, 3700Hz]. Note that the reverberation time, T_{60}, affects the variance of the direct arrival ATF, (4.41), from the setting $\bar{r} = 2r_c$ and the definition of the critical distance (4.45). In these figures, it is clearly depicted that the model for $\hat{\alpha}$, the variance of the reverberant component, is accurate, whereas the model for $\bar{\alpha}$, the variance of the direct arrival, demonstrates small mismatch. The model for α, the variance of the total ATF, is accurate, since it is mostly affected by the reverberant component. These results also apply when considering a specific frequency (in the specified range), instead of averaging over all frequencies.

In the third experiment, we verified our theoretical result stating that the ATFs are

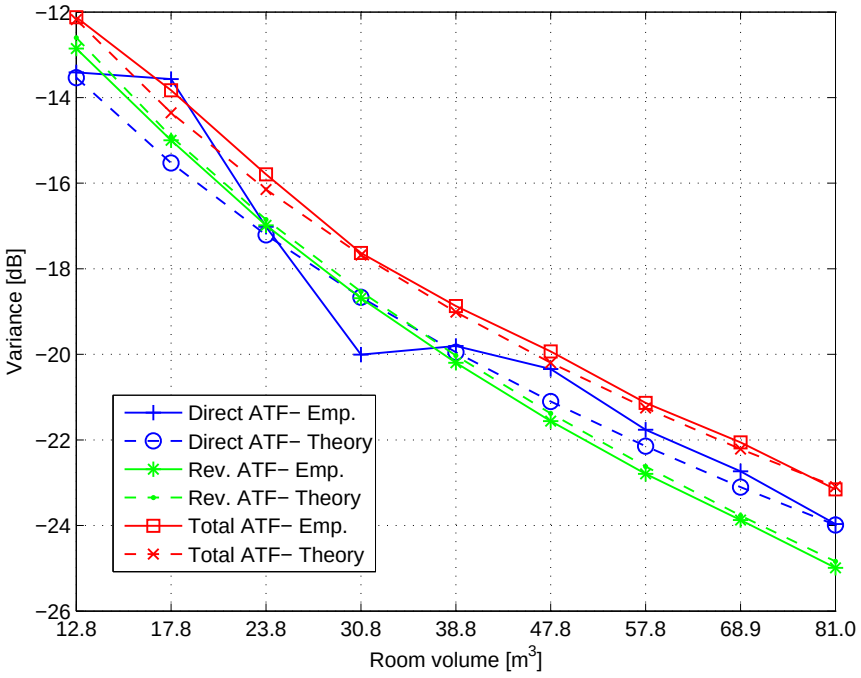


Figure 4.5: Empirical and theoretical ATF variances versus room dimensions.

uncorrelated. The room dimensions were set to 3m × 3m × 2.4m. We tested the statement for various reverberation times 0.2s, 0.4s and 0.6s, and for different distances between the sources $r_{1,2} \triangleq \|\boldsymbol{r}_1 - \boldsymbol{r}_2\| = 0.2\text{m}, 0.6\text{m}, \ldots, 1.8\text{m}$. For each room configuration 4 locations of the sources were randomly selected, and for each source location 100 different locations of microphones were randomized. The empirical coherence of the total ATFs is defined as $\text{coh}\,(h_1, h_2) \triangleq \frac{\langle h_1 h_2^* \rangle}{\sqrt{\langle |h_1|^2 \rangle \langle |h_2|^2 \rangle}}$, where we note that h_1 and h_2 are zero mean. In a similar manner, we define $\text{coh}\left(\bar{h}_1, \bar{h}_2\right)$, $\text{coh}\left(\hat{h}_1, \hat{h}_2\right)$. The empirical coherence of the direct ATFs, the reverberant ATFs and the total ATFs for all tested reverberation times and distances between sources are lower than −30dB (averaged over all frequencies). The empirical coherence versus frequency in the case of $T_{60} = 0.4\text{sec}$ and $r_{1,2} = 0.2\text{m}$ is depicted in Fig. 4.6. The results of this evidently verify the assumption that the ATFs of different sources are uncorrelated. In all experiments a sampling rate of 8kHz is used.

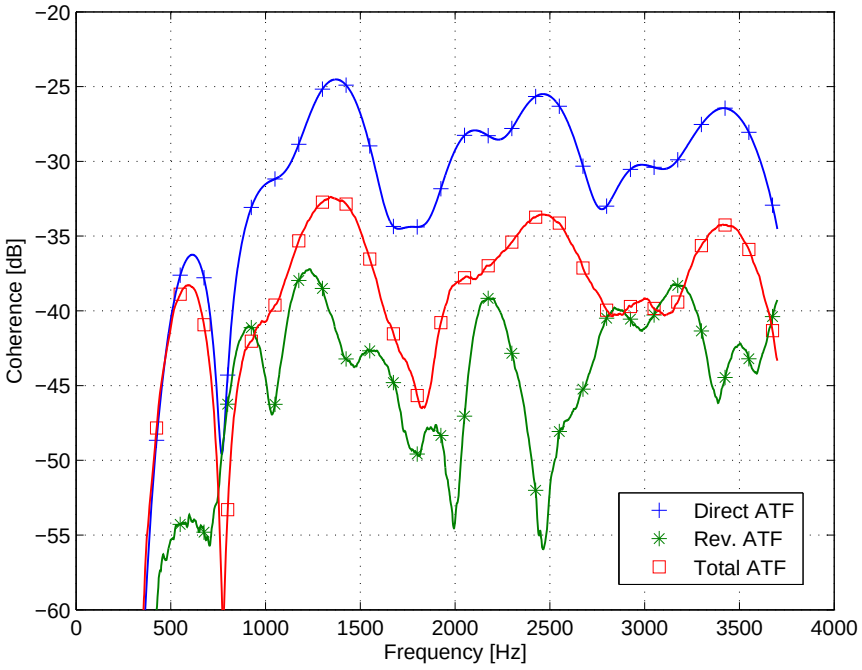


Figure 4.6: Empirical coherence between ATFs versus frequency.

4.2.4 Beamformers performance

In this section, we analyze the performance of the SDW-MWF in various noise fields. We derive *reliability* measures for the SIR and white noise gain criteria. The reliability of an SIR level of κ_0 is defined as the probability that the output SIR will exceed κ_0, explicitly:

$$R_{\kappa}\left(\kappa_0\right) \triangleq \Pr\left(\kappa \geq \kappa_0\right). \tag{4.53}$$

Similarly, the reliability of a white noise gain level of ξ_0, is defined as the probability that the white noise gain will exceed ξ_0:

$$R_{\xi}\left(\xi_0\right) \triangleq \Pr\left(\xi \geq \xi_0\right). \tag{4.54}$$

The reliability functions can be used to predict the performance of the BF in the WASN. Moreover, they can be used to determine the number of microphones that should be used in order to meet a predefined performance level. However, as these measures are statistical, for any microphones location realization, a non-zero probability that the desired performance

level will not be met exists.

4.2.4.1 Coherent interference signals $P < M$

Let $s_{i,1}(\ell), \ldots, s_{i,P}(\ell)$ be $P < M$ coherent noise sources located at $\boldsymbol{r}_{i,1}, \ldots, \boldsymbol{r}_{i,P}$, respectively. Denote the covariance matrix of the interference signals as:

$$\boldsymbol{\Delta}_i \triangleq \operatorname{diag}\{\delta_{i,1}, \ldots, \delta_{i,P}\} \tag{4.55}$$

where $\delta_{i,p} = \mathrm{E}\{|s_{i,p}(\ell)|^2\}$ is the variance of the pth source, for $p = 1, \ldots, P$, and let

$$\boldsymbol{h}_{i,p} \triangleq \begin{bmatrix} h_{i,p}^1 & \cdots & h_{i,p}^M \end{bmatrix}^T \tag{4.56}$$

be the ATF relating the pth interfering source and the microphones. The received interference signals vector, in the case of coherent interference signals, is therefore given by:

$$\boldsymbol{v}(\ell) = \boldsymbol{H}_i \boldsymbol{s}_i(\ell) + \boldsymbol{u}(\ell) \tag{4.57}$$

where

$$\boldsymbol{H}_i = \begin{bmatrix} \boldsymbol{h}_{i,1} & \cdots & \boldsymbol{h}_{i,P} \end{bmatrix} \tag{4.58a}$$

$$\boldsymbol{s}_i(\ell) = \begin{bmatrix} s_{i,1}(\ell) & \cdots & s_{i,P}(\ell) \end{bmatrix}^T \tag{4.58b}$$

and $\boldsymbol{u}(\ell)$ is a complex Gaussian sensors noise with covariance $\delta_u \boldsymbol{I}$, i.e. $\boldsymbol{u}(\ell) \sim \mathcal{CN}(\boldsymbol{0}, \delta_u \boldsymbol{I})$. Therefore, the covariance matrix of the received interference signals is:

$$\boldsymbol{\Phi}_{vv} = \boldsymbol{H}_i \boldsymbol{\Delta}_i \boldsymbol{H}_i^H + \delta_u \boldsymbol{I}. \tag{4.59}$$

Consider the expression $\boldsymbol{h}_d^H \boldsymbol{\Phi}_{vv}^{-1} \boldsymbol{h}_d$ which appears in the SIR and white noise gain criteria of the SDW-MWF in (4.30a) and (4.30b). Applying the Woodbury identity to $\boldsymbol{\Phi}_{vv}^{-1}$ as defined in (4.59) yields:

$$\begin{aligned} \boldsymbol{\Phi}_{vv}^{-1} =& \delta_u^{-1} \boldsymbol{I}_M - \delta_u^{-1} \boldsymbol{H}_i \left(\boldsymbol{I}_P + \delta_u \left(\boldsymbol{H}_i^H \boldsymbol{H}_i \right)^{-1} \boldsymbol{\Delta}_i^{-1} \right)^{-1} \\ & \times \left(\boldsymbol{H}_i^H \boldsymbol{H}_i \right)^{-1} \boldsymbol{H}_i^H. \end{aligned} \tag{4.60}$$

Now, assuming that the power of the coherent interference signals is much larger than the

variance of the sensors noise, and further assuming that $\boldsymbol{I}_P+\delta_u \left(\boldsymbol{H}_i^H \boldsymbol{H}_i\right)^{-1} \boldsymbol{\Delta}_i^{-1} \approx \boldsymbol{I}_P$, (4.60) can be approximated as:

$$\boldsymbol{\Phi}_{vv}^{-1} \approx \delta_u^{-1} \left(\boldsymbol{I}_M - \boldsymbol{H}_i \left(\boldsymbol{H}_i^H \boldsymbol{H}_i\right)^{-1} \boldsymbol{H}_i^H\right). \tag{4.61}$$

Note that, $\boldsymbol{\Phi}_{vv}^{-1}$ is approximately a projection matrix to the null-subspace of $\boldsymbol{H}_i$, scaled by δ_u^{-1}.

Let

$$\boldsymbol{H}_i = \boldsymbol{\Psi}\boldsymbol{\Omega}\boldsymbol{\Theta}^H \tag{4.62}$$

be the SVD of $\boldsymbol{H}_i$. Substituting (4.62) in (4.61), we obtain the more compact expression:

$$\boldsymbol{\Phi}_{vv}^{-1} = \delta_u^{-1} \dot{\boldsymbol{\Psi}} \dot{\boldsymbol{\Psi}}^H \tag{4.63}$$

where $\dot{\boldsymbol{\Psi}}$ is an $M \times (M-P)$ matrix comprising the $M-P$ last columns of $\boldsymbol{\Psi}$, associated with the zero singular values, which span the null-subspace of $\boldsymbol{H}_i$.

Define

$$\boldsymbol{\rho}_c \triangleq \dot{\boldsymbol{\Psi}}^H \boldsymbol{h}_d \tag{4.64}$$

and substituting (4.63) and (4.64) in (4.30a) and (4.30b) yields the simplified criteria expressions:

$$\kappa_c = \frac{\delta_d}{\delta_u} \|\boldsymbol{\rho}_c\|^2 \tag{4.65a}$$

$$\xi_c = \|\boldsymbol{\rho}_c\|^2 \tag{4.65b}$$

where we denote the SIR and white noise gain criteria in the coherent noise field (for $P < M$) as κ_c and ξ_c, respectively. Note that both criteria depend on $\|\boldsymbol{\rho}_c\|^2$ with different coefficient multipliers. Now, we turn to analyze their statistics.

Denote the nth column of $\dot{\boldsymbol{\Psi}}$ and the nth element of $\boldsymbol{\rho}_c$ as $\dot{\boldsymbol{\psi}}_n$ and $\rho_{c,n}$, respectively, for $n = 1, \ldots, (M-P)$. A single element $\rho_{c,n}$ is obtained by

$$\rho_{c,n} = \dot{\boldsymbol{\psi}}_n^H \boldsymbol{h}_d \tag{4.66}$$

which is a linear combination of the uncorrelated elements of $\boldsymbol{h}_d$. From the unitarity of $\boldsymbol{\Psi}$, we conclude that $\boldsymbol{\rho}_c$ is a vector of uncorrelated variables:

$$\begin{aligned}\mathrm{E}\left\{\boldsymbol{\rho}_c\boldsymbol{\rho}_c^H\right\} =&\mathrm{E}\left\{\dot{\boldsymbol{\Psi}}_i^H\boldsymbol{h}_d\boldsymbol{h}_d^H\dot{\boldsymbol{\Psi}}_i\right\}\\ =&\alpha\boldsymbol{I}_{M-P}.\end{aligned} \tag{4.67}$$

Now, since $(M-P)\gg 1$ the CLT conditions hold and hence we argue that the distribution of the random variable $\rho_{c,n};\ n=1,\ldots,M-P$ converges to the complex normal distribution $\rho_{c,n}\sim\mathcal{CN}(0,\alpha)$, where α is defined in (4.47):

$$\boldsymbol{\rho}_c\sim\mathcal{CN}\left(\mathbf{0},\alpha\boldsymbol{I}_{M-P}\right). \tag{4.68}$$

It is easily concluded that the elements of $\boldsymbol{\rho}_c$ are i.i.d..

Define

$$\eta_c\triangleq\frac{2}{\alpha}\|\boldsymbol{\rho}_c\|^2. \tag{4.69}$$

It is a Chi-square RV with $2(M-P)$ degrees of freedom, i.e.

$$\eta_c\sim\chi^2\left(2\left(M-P\right)\right). \tag{4.70}$$

From (4.69) we have:

$$\|\boldsymbol{\rho}_c\|^2=\frac{\alpha}{2}\eta_c. \tag{4.71}$$

Substituting (4.71) in (4.65a) and (4.65b) yields alternative expressions for the performance criteria:

$$\kappa_c=\frac{\delta_d}{\delta_u}\frac{\alpha}{2}\eta_c \tag{4.72a}$$

$$\xi_c=\frac{\alpha}{2}\eta_c. \tag{4.72b}$$

Using the probability distribution function (p.d.f.) of η_c, the average SIR and white noise

gain, denoted $\bar{\kappa}_c$ and $\bar{\xi}_c$, are given by:

$$\bar{\kappa}_c \triangleq \mathrm{E}\left\{\kappa_c\right\} = \frac{\alpha\delta_d}{\delta_u}\left(M-P\right) \tag{4.73a}$$

$$\bar{\xi}_c \triangleq \mathrm{E}\left\{\xi_c\right\} = \alpha\left(M-P\right) \tag{4.73b}$$

where we substituted $\mathrm{E}\left\{\eta_c\right\} = 2\left(M-P\right)$. Note, that the latter averages of the SIR and white noise gain are linear with the number of microphones.

From (4.72a) it is clear that κ_c is a scaled version of a Chi-square RV with $2(M-P)$ degrees of freedom. Hence, its reliability (4.53) can be calculated as:

$$R_{\kappa,c}\left(\kappa_0\right) = 1 - F_{\eta,c}\left(\frac{2}{\alpha}\frac{\delta_u}{\delta_d}\kappa_0\right) \tag{4.74}$$

where

$$\begin{aligned} F_{\eta,c}\left(\eta_0\right) \triangleq & \Pr\left(\eta_c \leq \eta_0\right) \\ = & \frac{\gamma_f\left(M-P, \frac{\eta_0}{2}\right)}{\Gamma_f\left(M-P\right)} \end{aligned} \tag{4.75}$$

is the cumulative distribution function of a Chi-square RV with $2\left(M-P\right)$ degrees of freedom, Γ_f is the Gamma function and γ_f is the lower incomplete Gamma function.

Similarly, the reliability of the white noise gain (4.72b) is:

$$R_{\xi,c}\left(\xi_0\right) = 1 - F_{\eta,c}\left(\frac{2}{\alpha}\xi_0\right). \tag{4.76}$$

4.2.4.2 Diffuse sound field

In this section we derive the performance of an SDW-MWF in a diffuse sound field. This noise field is can be modeled by numerous statistically independent noise sources arriving from all directions simultaneously ($P \gg M$). It is a common noise field in reverberant environments, cocktail party and car scenarios [117]. The covariance $\Phi_{vv}\left(m, m'\right)$ of a diffused noise between the noise components received at the mth and the m'th microphones equals:

$$\Phi_{vv}\left(m, m'\right) = \delta_{\mathrm{dif}}\,\mathrm{sinc}\left(\frac{2\pi\|\boldsymbol{r}^m - \boldsymbol{r}^{m'}\|}{\lambda_k}\right) \tag{4.77}$$

where δ_{dif} denotes the variance of the diffuse sound field received at each microphone. Note that the coherence is a RV due to the random microphone locations.

Consider, the average covariance $\mathrm{E}\left\{\Phi_{vv}\left(m,m'\right)\right\}$ in a sphere with radius $R \gg \lambda_k$ around $\boldsymbol{r}^{m'}$. Assuming that the mth microphone is randomly located in the sphere with a uniform distribution, the expectation can be formulated as:

$$\begin{aligned}\mathrm{E}\left\{\Phi_{vv}\left(m,m'\right)\right\} =& \iiint \frac{\delta_{\text{dif}}\,\mathrm{sinc}\left(\frac{2\pi r}{\lambda_k}\right) r\sin\left(\theta\right)\mathrm{d}r\mathrm{d}\theta\mathrm{d}\phi}{4/3\pi R^3} \\ =& \delta_{\text{dif}} \times \begin{cases} 1 & ;\ m = m' \\ \frac{3\sin^2\left(\frac{\pi R}{\lambda_k}\right)}{2R} & ;\ m \neq m' \end{cases}. \end{aligned} \tag{4.78}$$

Now, since $R \gg \lambda_k$ we can approximate:

$$\mathrm{E}\left\{\boldsymbol{\Phi}_{vv}\left(m,m'\right)\right\} \approx \delta_{\text{dif}}. \tag{4.79}$$

Since the enclosure is much larger than λ_k, we assume that on average the distance between any pair of microphones is larger than λ_k, and propose the approximation:

$$\boldsymbol{\Phi}_{vv} \approx \delta_{\text{dif}}\boldsymbol{I}. \tag{4.80}$$

Define

$$\boldsymbol{\rho}_{\text{dif}} \triangleq \sqrt{\delta_{\text{dif}}}\left(\boldsymbol{\Phi}_{vv}^{-1/2}\right)^H \boldsymbol{h}_d \tag{4.81}$$

where $\boldsymbol{\Phi}_{vv}^{-1} = \boldsymbol{\Phi}_{vv}^{-1/2}\left(\boldsymbol{\Phi}_{vv}^{-1/2}\right)^H$ is the Cholesky decomposition. Since, in most cases the power of the reverberant component dominates the ATF, we propose to model $\boldsymbol{h}_d \sim \mathcal{CN}\left(\boldsymbol{0},\alpha\boldsymbol{I}_{M\times M}\right)$, and by using (4.80) to model $\boldsymbol{\rho}_{\text{dif}}$ as an $M \times 1$ complex Gaussian RV with the probability distribution:

$$\boldsymbol{\rho}_{\text{dif}} \sim \mathcal{CN}\left(\boldsymbol{0},\alpha\boldsymbol{I}_{M\times M}\right). \tag{4.82}$$

Define

$$\eta_{\text{dif}} \triangleq \frac{2}{\alpha}\|\boldsymbol{\rho}_{\text{dif}}\|^2 \tag{4.83}$$

and note that η_{dif} is a Chi-square RV with $2M$ degrees of freedom, i.e.

$$\eta_{\text{dif}} \sim \chi^2\left(2M\right). \tag{4.84}$$

Substituting (4.80), (4.81) and (4.83) in (4.30a) and (4.30b) yields:

$$\kappa_{\text{dif}} = \frac{\delta_d}{\delta_{\text{dif}}} \frac{\alpha}{2} \eta_{\text{dif}} \tag{4.85a}$$

$$\xi_{\text{dif}} = \frac{\alpha}{2} \eta_{\text{dif}} \tag{4.85b}$$

where we have applied the following approximation resulting from (4.80):

$$\mathbf{\Phi}_{vv}^{-2} \approx \frac{1}{\delta_{\text{dif}}} \mathbf{\Phi}_{vv}^{-1}. \tag{4.86}$$

The average SIR and white noise gain in the diffuse sound field case are given by:

$$\bar{\kappa}_{\text{dif}} \triangleq \mathrm{E}\left\{\kappa_{\text{dif}}\right\} = \frac{\alpha \delta_d}{\delta_{\text{dif}}} M \tag{4.87a}$$

$$\bar{\xi}_{\text{dif}} \triangleq \mathrm{E}\left\{\xi_{\text{dif}}\right\} = \alpha M \tag{4.87b}$$

where we substitute $\mathrm{E}\left\{\eta_{\text{dif}}\right\} = 2M$. Note, that as in the coherent interference signals case, the latter averages of the SIR and white noise gain are linear with the number of microphones. Van Trees [118] showed that, for spatially white noise, the SIR linearly increases with the number of microphones in the deterministic case (when the microphone locations are not random). Since we show that the diffuse sound field covariance matrix can be approximated by a scaled identity matrix, we obtain a similar result for randomly located microphones.

Similarly to (4.74), (4.76) the reliability of κ_{dif} and ξ_{dif} are given by:

$$R_{\kappa,\text{dif}}\left(\kappa_0\right) = 1 - F_{\eta,\text{dif}}\left(\frac{2}{\alpha} \frac{\delta_{\text{dif}}}{\delta_d} \kappa_0\right) \tag{4.88a}$$

$$R_{\xi,\text{dif}}\left(\xi_0\right) = 1 - F_{\eta,\text{dif}}\left(\frac{2}{\alpha} \xi_0\right) \tag{4.88b}$$

where

$$\begin{aligned} F_{\eta,\text{dif}}\left(\eta_0\right) &\triangleq \Pr\left(\eta_{\text{dif}} \leq \eta_0\right) \\ &= \frac{\gamma_f\left(M, \frac{\eta_0}{2}\right)}{\Gamma_f\left(M\right)} \end{aligned} \tag{4.89}$$

is the cumulative distribution function of the $2M$ degrees of freedom Chi-square RV η_{dif}.

4.2.5 BF Model verification

4.2.5.1 Coherent interference signals $P < M$

We carry out experiments to verify the theoretical model for the statistics of κ_c and ξ_c in the case of coherent noise sources. The room dimensions are set to 4m $\times$ 4m $\times$ 3m, the number of microphones is $M = 5$, the sampling rate is 8kHz, and the DFT size is $K = 8192$. The number of interference signals is in the range of $P = 1, 2, 3, 4$ and the reverberation time can take the values $0.2\text{s}, 0.4\text{s}, 0.6\text{s}, 0.8\text{s}$. The received microphone signals comprise of a coherent desired source, modeled as a 6th order auto regressive (AR) random process, coherent noise sources, modeled as an AR(1) random processes, and sensors noise. We simulate the spectra of the signals, and substitute them in the derived formulas. The average SNR of the desired source, and the average INR of each of the coherent noise sources are set to 90dB at the microphones. The locations of the desired source and the interference signals are randomly selected in 4 scenarios. In each scenario 100 microphones positions are drawn with a 3D uniform distribution. In each Monte-Carlo experiment, and per each frequency, the SDW-MWF with $\mu = 1$ is calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e. κ_c and ξ_c, defined as $\frac{(\langle\kappa_c\rangle-\bar{\kappa}_c)^2}{\bar{\kappa}_c^2}$ and $\frac{\left(\langle\xi_c\rangle-\bar{\xi}_c\right)^2}{\bar{\xi}_c^2}$, respectively, are -20dB for low and medium reverberation times, $0.2\text{s}, 0.4\text{s}$, and for all numbers of interference signals scenarios. For higher reverberation times, $0.6\text{s}, 0.8\text{s}$, the measured normalized errors are a bit higher, -15dB. Evidently, the formulas for the average criteria (4.73a),(4.73b) are valid.

The following figures correspond to one of the desired source and interference signals constellations at frequency 2kHz. Similar results are obtained for other scenarios and frequencies. In the derivation of the theoretical model, we argued that κ_c is a scaled $\chi^2\left(2\left(M-P\right)\right)$ RV. The quantile-quantile probability plots of $\frac{2\delta_u}{\alpha\delta_d}\kappa_c$ versus the $\chi^2\left(2\left(M-P\right)\right)$ distribution is depicted in Fig. 4.7 for reverberation time $T_{60} = 0.4\text{s}$, and for various numbers of interference signals $P = 1, \ldots, 4$. From this figure, the Chi-square distribution with $2\left(M-P\right)$ degrees of freedom of the scaled κ_c can be verified. We also verify that $\boldsymbol{\rho}_c$ is an $(M-P)\times 1$ complex normal vector. The reliability function of the SIR, i.e. $R_{\kappa,c}$, versus the SIR improvement (defined as the ratio of the output and input SIR) for $T_{60} = 0.4\text{s}$ is depicted in Fig. 4.8. Clearly from this figure, the reliability function of the SIR is verified. As expected, the reliability of the white noise gain demonstrates similar behavior.

The reliability functions of the SIR and white noise gain were measured for all combi-

nations of $T_{60} = 0.2, 0.4, 0.6, 0.8$s and $P = 1, 2, 3, 4$ interference signals for various values of $\mu = 1, 10, 100$. Correspondingly to derivation in (4.74), (4.76), the measured reliability criteria are independent to the parameter μ.

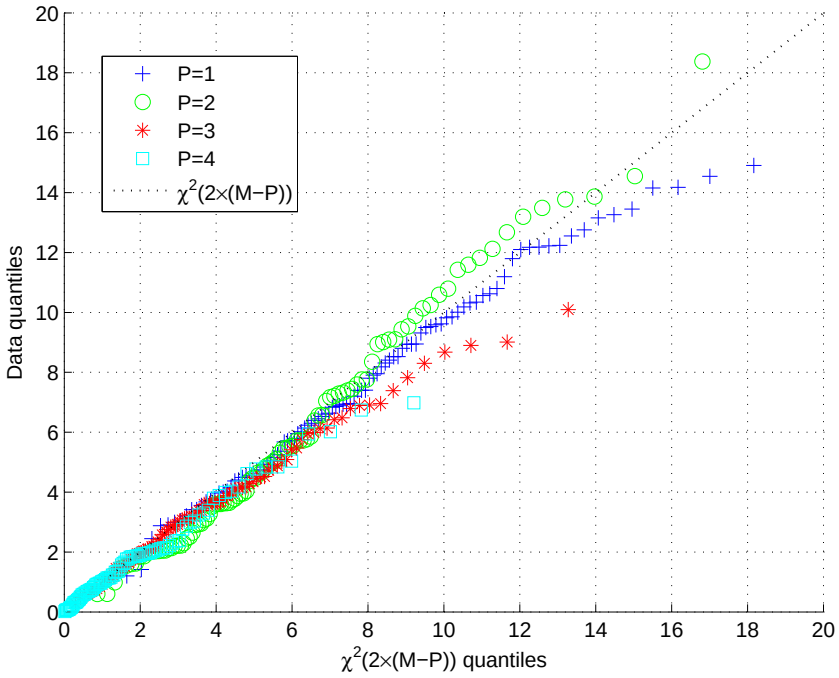


Figure 4.7: Quantile-quantile probability plot of $\frac{2\delta_u}{\alpha\delta_d}\kappa_c$ versus $\chi^2\left(2\left(M-P\right)\right)$ distribution with various numbers of coherent interference signals, $P = 1, \ldots, 4$, and for a reverberation time of $T_{60} = 0.4$s.

Now, we wish to verify the effect of the number of microphones M on the reliability measures and of the SIR and white noise gain in the coherent interference signals case. We use the same room dimensions as above, and set the reverberation time to $T_{60} = 0.4$s. We test 4 different constellations of a desired source and a single interference signal. For each constellation 100 microphone locations are uniformly randomized, where the number of microphones is taken from $M = 5, 10, 15, 20, 25$. As before, in each Monte-Carlo experiment and per each frequency, the SDW-MWF with $\mu = 1$ is calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e. κ_c and ξ_c, are -20dB for all tested numbers of microphones. The theoretical relation between the number of microphones, M, and the average SIR and white noise gain (4.73a),(4.73b), are verified, as the normalized errors are considerably small.

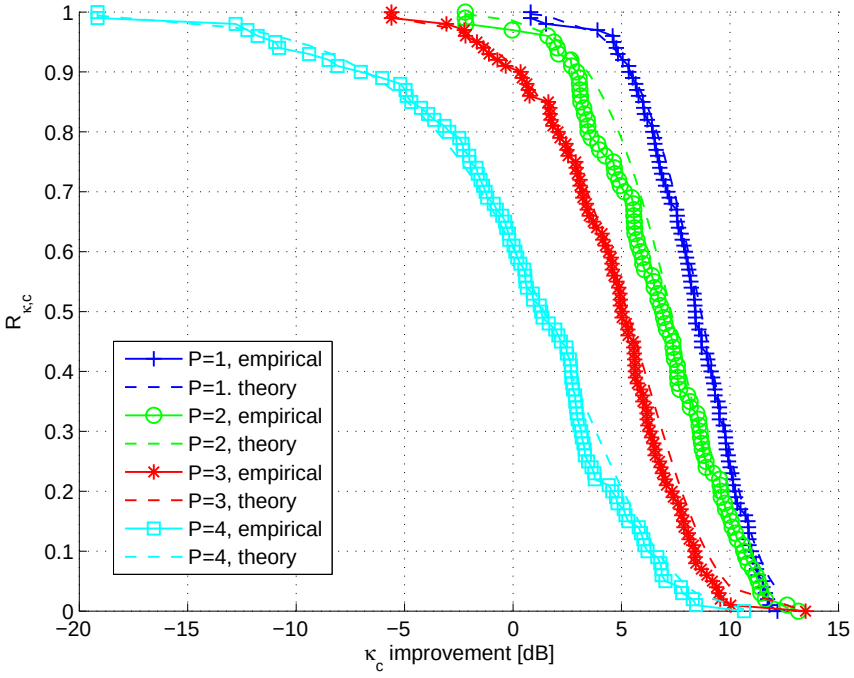


Figure 4.8: Reliability function of SIR, $R_{\kappa,c}$, versus SIR improvement with various numbers of coherent interference signals, $P = 1, \ldots, 4$, and for a reverberation time of $T_{60} = 0.4$s.

The reliability of the SIR at a frequency of 2kHz, i.e. κ_c versus the SIR improvement for various numbers of microphones is depicted in Fig. 4.9. It is clear from this figure, that the derived reliability function fits the empirical data. It is interesting to note that as the number of microphones increases the reliability function converges to a step function, and hence the performance level becomes more deterministic. Similar results are obtained for other frequencies and sources constellations. As discussed earlier, the reliability measures (4.88a),(4.88b) equal the probability that the performance criteria will meet a predefined level.

4.2.5.2 Diffuse sound field

Here, we perform an experiment to verify the theoretical model of κ_{dif} and ξ_{dif} for the case of a diffuse sound field. The room dimensions, the sampling rate and the DFT size are as in the previous section, 4m $\times$ 4m $\times$ 3m, 8kHz and 8192, respectively. The number of microphones is set to $M = 16$. The reverberation time is set to one of the values $0.2\text{s}, 0.3\text{s}, \ldots, 0.6\text{s}$. The received microphone signals comprise of a coherent desired source, modeled as before by an AR(6) random process, a diffuse sound field and sensors noise. As in the previous experiment,

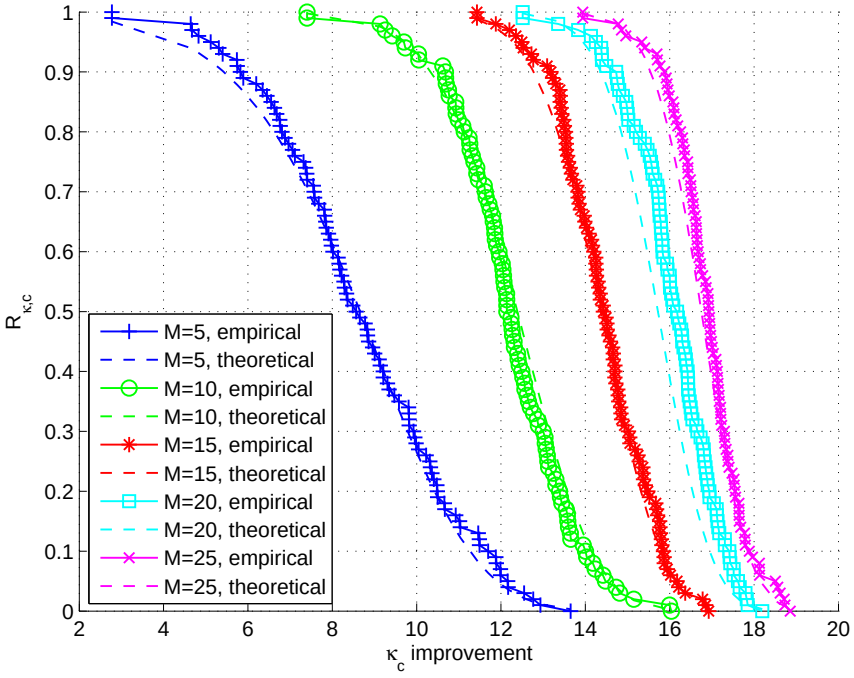


Figure 4.9: Reliability function of SIR, $R_{\kappa,c}$, versus SIR improvement with various numbers of microphones.

we simulate the spectra of the signals, and substitute them in the derived formulas. The average SNR of the desired source and the average INR of the diffuse sound field are set to 90dB and 60dB, respectively. The location of the desired source is randomly selected in 4 scenarios. For each scenario, 100 microphones positions are drawn with a 3D uniform distribution. In each Monte-Carlo experiment, and per each frequency, the SDW-MWF with $\mu = 1$ is calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e. κ_{dif} and ξ_{dif}, defined as $\frac{(\langle\kappa_{\text{dif}}\rangle-\bar{\kappa}_{\text{dif}})^2}{\bar{\kappa}_{\text{dif}}^2}$ and $\frac{\left(\langle\xi_{\text{dif}}\rangle-\bar{\xi}_{\text{dif}}\right)^2}{\bar{\xi}_{\text{dif}}^2}$, respectively, for all tested reverberation times is about -20dB. Evidently, the formulas for the average criteria (4.87a),(4.87b) are valid.

The following figures correspond to one of the source location scenarios at frequency 2kHz, however, similar results are obtained at other scenarios and frequencies. In the derivation of the theoretical model, we argued that κ_{dif} is a scaled $\chi^2(2M)$ RV. The quantile-quantile probability plots of $\frac{2\delta_{\text{dif}}}{\alpha\delta_d}\kappa_{\text{dif}}$ versus the $\chi^2(32)$ distribution is depicted in Fig. 4.10 for reverberation times of $0.2\text{sec}, 0.6\text{sec}$. From this figure, the Chi-square distribution with $2M$ degrees of freedom of the scaled κ_{dif} can be verified. The theoretical model is verified also for other

reverberation times. We also verify that $\boldsymbol{\rho}_{\text{dif}}$ is an $M \times 1$ complex normal random vector. The reliability function of the SIR, i.e. $R_{\kappa,\text{dif}}$, versus the SIR improvement is depicted in Fig. 4.11. Clearly from this figure, the reliability function of the SIR is verified. Similar results were obtained in all other tested reverberation times. The theoretical model for the reliability of the white noise gain is also verified in this simulation.

The reliability functions of the SIR and white noise gain, i.e. $R_{c,\text{dif}}$ and $R_{\xi,\text{dif}}$, were measured with different reverberation times $T_{60} = 0.2, 0.4, 0.6, 0.8$s and for various values of $\mu = 1, 10, 100$. Correspondingly to derivation in (4.88a), (4.88b), the reliability criteria are independent to the parameter μ.

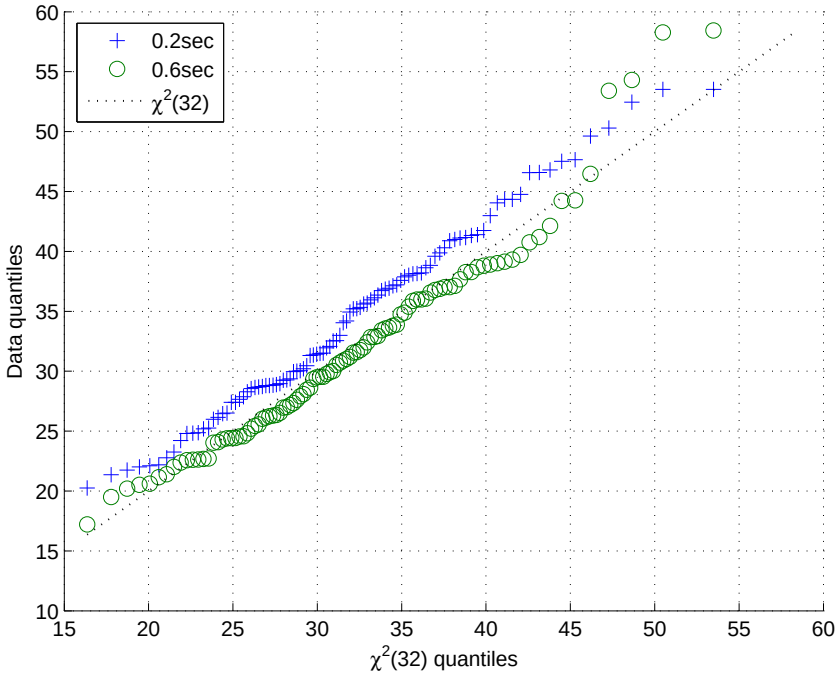


Figure 4.10: Quantile-quantile probability plot of $\frac{2\delta_{\text{dif}}}{\alpha\delta_d}\kappa_{\text{dif}}$ versus $\chi^2(32)$ distribution with a diffuse sound field for various reverberation times.

Now, we wish to verify the effect of the number of microphones M on the reliability measures of the SIR and white noise gain in the diffuse sound field case. We use the same room dimensions as above, and set the reverberation time to $T_{60} = 0.4$s. We test 4 different locations for the desired source. For each case 100 microphones locations are uniformly randomized, where the number of microphones is taken from $M = 5, 10, 15, 20, 25$. As before, in each Monte-Carlo experiment and per each frequency, the SDW-MWF with $\mu = 1$ is

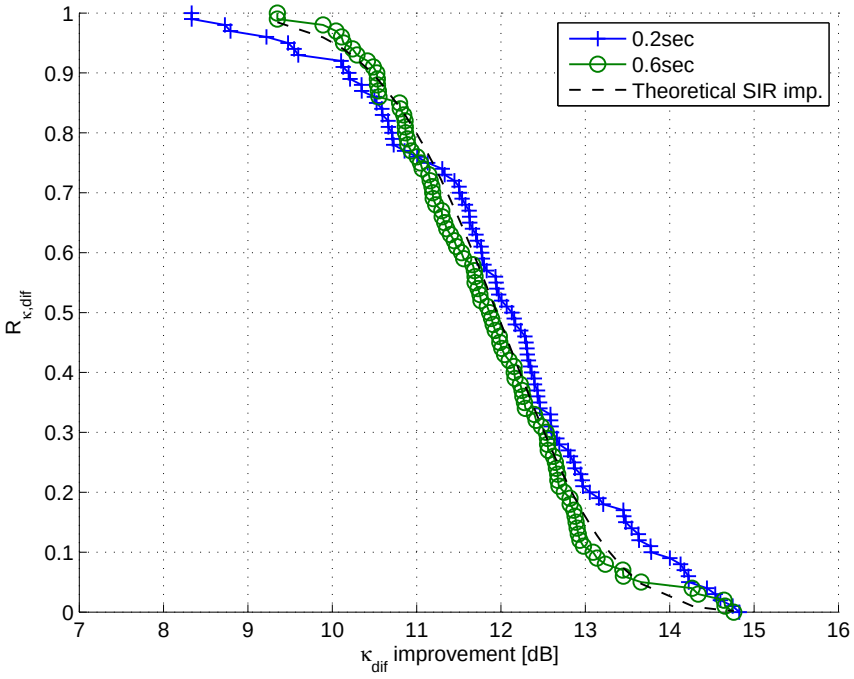


Figure 4.11: Reliability function of SIR, $R_{\kappa,\mathrm{dif}}$, versus SIR improvement with a diffuse sound field for various reverberation times.

calculated, and its SIR and white noise gain are recorded. The normalized errors of the average SIR and white noise gain, i.e. κ_{dif} and ξ_{dif}, for all tested numbers of microphones, are about -20dB. The theoretical relation between the number of microphones, M, and the average SIR and white noise gain (4.87a),(4.87b), are verified from these results, as the normalized errors are considerably small.

The reliability of the SIR at a frequency of 2kHz, i.e. κ_{dif}, versus the SIR improvement for various numbers of microphones is depicted in Fig. 4.12. Clearly from this figure, the derived reliability function fits the empirical data. Similar results are obtained for other frequencies and source locations. As in the case of coherent noise sources, the performance tends to become deterministic as the number of microphones increases.

4.2.6 Conclusions

We have considered the problem of signal enhancement in WASN applications where the microphone locations cannot be determined in advance. Assuming that the microphones are randomly located with a uniform distribution, and utilizing results from statistical room

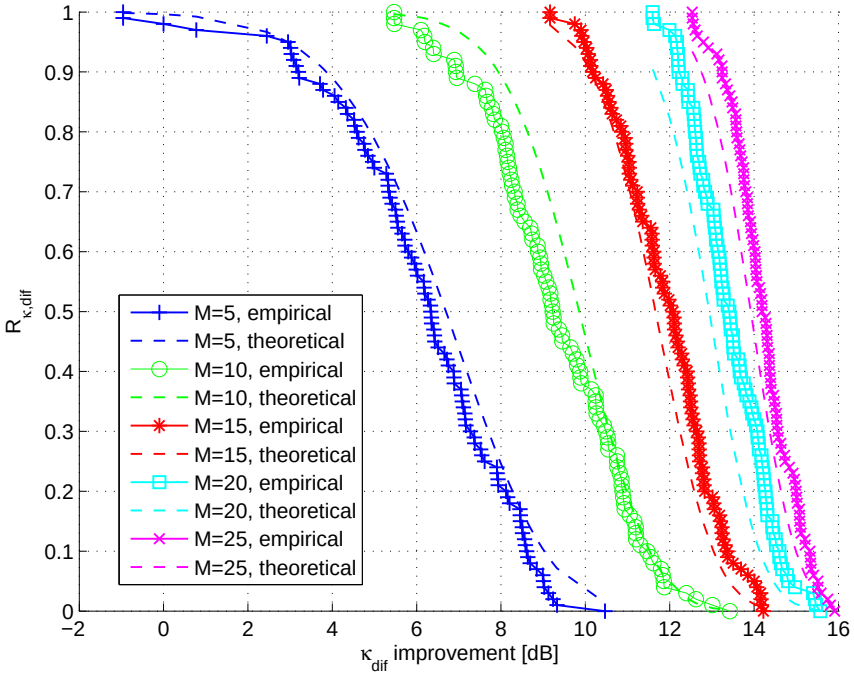


Figure 4.12: Reliability function of SIR, $R_{\kappa,\mathrm{dif}}$, versus SIR improvement with various numbers of microphones.

acoustics, we analyzed the performance of applying the SDW-MWF. Two noise fields were discussed: first, $P < M$ coherent noise sources and second, a diffuse sound field. Statistical models for two performance criteria, namely the SIR and the white noise gain, were derived for the different noise fields. Reliability functions, which give the probability of a BF criterion to exceed a predefined level, were derived for both criteria and both noise fields. The reliability functions can be used to predict the BF performance measures in a WASN, and to calculate the number of microphones needed to maintain a desired level thereof with a predefined probability. The proposed statistical models and reliability functions were verified in a comprehensive simulative study.

Chapter 5

Complexity reduced beamformers

The processing power of WASN is a restricting requirement for distributed algorithms. In this chapter we derive computationally efficient procedures for implementing a distributed LCMV-BF. In Sec. 5.1, we consider reducing the complexity of the GSC-BF, which is dominated by the BM filtering operation. We propose a sparse BM structure which significantly reduces the complexity involved in applying this stage. We show that this structure is equivalent to other BM implementations, and does not sacrifice the performance of the GSC-BF.

The dynamics of the WASN, manifested in variations of the network connectivity and in the activity patterns of speakers, is addressed in Sec. 5.2. We derive, efficient procedures for updating either the LCMV or GSC BFs correspondingly to changes in nodes connectivity and sources activity. These procedures reduce the computations requirement significantly compared with straightforward methods.

5.1 A sparse blocking matrix for multiple constraints GSC beamformer

The major contributor for computational burden in GSC-BF implementations is the BM. In this section a novel systematic procedure for designing a K constraints sparse $M \times (M-K)$ BM is proposed. The BM requires only $K \times (M-K)$ complex multiplications. The blocking ability of the sparse BM, defined as the robustness to the ATFs, is analyzed and compared with the blocking ability of the commonly used eigen-space BM. For low estimation errors, it is proven that blocking ability of the sparse BM and of the eigen-space BM are equivalent.

The section is organized as follows. In Sec. 5.1.1 the problem is formulated. In Sec. 5.1.2

the eigen-space BM and the sparse BM are formally derived. Then, in Sec. 5.1.3 the blocking ability and the signal leakage of the BMs are analyzed. A comprehensive experimental study of narrowband signals as well as of speech signals is described in Sec. 5.1.4.

5.1.1 Problem Formulation

Consider a microphone array comprising M microphones. The received signals in the STFT domain are:

$$\boldsymbol{z}(\ell,\zeta) = \boldsymbol{H}(\zeta)\boldsymbol{s}(\ell,\zeta) + \boldsymbol{v}(\ell,\zeta) \tag{5.1}$$

where ℓ is the frame index and ζ is the frequency bin index. The received signals comprise two contributions. The first contribution $\boldsymbol{H}(\zeta)\boldsymbol{s}(\ell,\zeta)$ is related to the constrained sources, where $\boldsymbol{s}(\ell,\zeta) = \begin{bmatrix} s_1(\ell,\zeta) & \cdots & s_K(\ell,\zeta) \end{bmatrix}^T$ is a $K \times 1$ vector of coherent signals, $\boldsymbol{H}(\zeta) = \begin{bmatrix} \boldsymbol{h}_1(\zeta) & \cdots & \boldsymbol{h}_K(\zeta) \end{bmatrix}$ is an $M \times K$ constraints matrix comprised of the K ATFs relating the constrained sources and the microphones. The second contribution, $\boldsymbol{v}(\ell,\zeta)$, is related to the non-constrained contribution. Without loss of generality, we assume that the ATFs are normalized, i.e. $\|\boldsymbol{h}_k(\zeta)\|^2 = 1;\ k = 1,..,K$. Henceforth, the frequency bin index ζ is omitted for brevity. The derived formulas correspond to either a single frequency, or one frequency bin for wideband signals.

5.1.2 Designing the BM

In Sec. 5.1.2.1 the eigen-space based BM is defined, as in [51]. In Sec. 5.1.2.2 the proposed sparse BM is derived.

5.1.2.1 Eigen-space based BM

The eigen-space BM is given by the projection matrix to the null-subspace of the constraint matrix $\boldsymbol{H}$:

$$\boldsymbol{B}_e = \boldsymbol{I}_{M\times M} - \boldsymbol{H}\left(\boldsymbol{H}^H\boldsymbol{H}\right)^{-1}\boldsymbol{H}^H \tag{5.2}$$

where $\boldsymbol{I}_{M\times M}$ is the $M \times M$ identity matrix. It can be verified that $\boldsymbol{B}_e^H\boldsymbol{H} = \boldsymbol{0}$. Application of the eigen-space BM involves M^2 complex multiplications per frame and frequency bin.

5.1.2.2 Sparse BM

The contribution of the constrained signals to the received signals in (5.1), i.e. $\boldsymbol{H}\boldsymbol{s}(\ell)$, lies in a rank-$K$ subspace in the M dimensional space. By a proper transformation, $\boldsymbol{H}\boldsymbol{s}(\ell)$ can be expressed as a linear combination of the constrained signal contributions of K reference received signals. Without loss of generality, we consider the first K microphones as the reference signals. Denote the reference microphones by $\boldsymbol{z}_r(\ell) = \begin{bmatrix} z_1(\ell) & \cdots & z_K(\ell) \end{bmatrix}^T$. The reference microphones are given by:

$$\boldsymbol{z}_r(\ell) = \boldsymbol{H}_r\boldsymbol{s}(\ell) + \boldsymbol{v}_r(\ell) \tag{5.3}$$

where $\boldsymbol{H}_r = \boldsymbol{H}_{1:K,1:K}$, and $\boldsymbol{v}_r(\ell) = \begin{bmatrix} v_1(\ell) & \cdots & v_K(\ell) \end{bmatrix}^T$. Assuming that $\boldsymbol{H}_r$ is invertible, $\boldsymbol{H}\boldsymbol{s}(\ell)$ can be expressed in terms of $\boldsymbol{H}_r\boldsymbol{s}(\ell)$ as $\boldsymbol{H}\boldsymbol{s}(\ell) = \begin{bmatrix} \boldsymbol{I}_{K\times K} & \boldsymbol{\beta}_{K+1} & \cdots & \boldsymbol{\beta}_M \end{bmatrix}^H \boldsymbol{H}_r\boldsymbol{s}(\ell)$ where $\boldsymbol{\beta}_m = \left(\boldsymbol{H}_r^{-1}\right)^H \boldsymbol{H}_{m,:}^H$ for $m = K+1,..,M$ and $\boldsymbol{H}_{m,:}$ is the mth row of $\boldsymbol{H}$. Utilizing the latter representation, a noise reference (non-constrained part) based on the mth microphone (for $m = K+1,..,M$), is extracted by subtracting a linear combination of the reference microphones $\boldsymbol{z}_r(\ell)$ from $z_m(\ell)$. I.e.

$$u_m(\ell) = z_m(\ell) - \boldsymbol{\beta}_m^H\boldsymbol{z}_r(\ell) = v_m(\ell) - \boldsymbol{\beta}_m^H\boldsymbol{v}_r(\ell). \tag{5.4}$$

The corresponding BM is denoted as $\boldsymbol{B}_s$ and is given by:

$$\boldsymbol{B}_s = \begin{bmatrix} -\boldsymbol{\beta}_{K+1} & \cdots & -\boldsymbol{\beta}_M \\ & \boldsymbol{I}_{(M-K)\times(M-K)} & \end{bmatrix}. \tag{5.5}$$

Please note that $\boldsymbol{B}_s$ has $(M-K)\times K$ non-zero entries in its first K rows and $M-K$ entries equal to 1 in the lower $M-K$ rows. Hence, the proposed BM can be denoted as the sparse BM. Its application requires $(M-K)\times K$ complex multiplications per frame per frequency, which is much lower than the $M\times(M-K)$ complex multiplications required by the eigen-space BM (assuming that $K \ll M$).

In the special case of $K = 1$ the proposed sparse BM equals to the BM proposed by Gannot et al. [39], which is based on the RTF with respect to a single (arbitrarily chosen) microphone.

5.1.3 Performance analysis

In this section, the blocking ability and the signal leakage criteria are defined and analyzed for the eigen-space and for the sparse BMs.

Consider a noisy estimate of $\boldsymbol{H}$:

$$\tilde{\boldsymbol{H}} = \boldsymbol{H} + \boldsymbol{\Delta} \tag{5.6}$$

where $\boldsymbol{\Delta} = \begin{bmatrix} \boldsymbol{\delta}_1 & \cdots & \boldsymbol{\delta}_K \end{bmatrix}$ comprises the $M \times 1$ dimensional vectors $\boldsymbol{\delta}_1, \ldots, \boldsymbol{\delta}_K$ of i.i.d. complex Normal random variables with a zero mean, and a variance of λ_u. Since $\boldsymbol{h}_k$; $k = 1, .., K$ are assumed to be normalized, the estimation accuracy defined as $\|\boldsymbol{H}\|_{\mathrm{F}}^2 / \mathrm{E}\{\|\boldsymbol{\Delta}\|_{\mathrm{F}}^2\}$ equals $(M\lambda_u)^{-1}$, where $\|\cdot\|_{\mathrm{F}}^2$ is the squared Frobenius norm.

The ability of the noisy BM $\tilde{\boldsymbol{B}}_b$ to block $\boldsymbol{h}_k$, the ATF of the kth source, is denoted by η_b^k, and equals the ratio between the leakage of kth ATF to the output of the BM, $\lambda_b^{s,k}$, and the power of a unit variance spatially white noise filtered by the BM, λ_b^n:

$$\eta_b^k = \frac{\lambda_b^{s,k}}{\lambda_b^n} \tag{5.7}$$

where

$$\lambda_b^{s,k} = \mathrm{E}\left\{ \|\tilde{\boldsymbol{B}}_b^H \boldsymbol{h}_k\|^2 \right\} \tag{5.8a}$$

$$\lambda_b^{n} = \mathrm{E}\left\{ \|\tilde{\boldsymbol{B}}_b^H \boldsymbol{w}\|^2 \right\} \tag{5.8b}$$

$b \in \{e, s\}$ stands for sparse BM (s), or eigen-space BM (e), and $\boldsymbol{w}$ is an $M \times 1$ vector of zero mean, and unit variance complex Normal i.i.d. RVs. Substituting $\boldsymbol{h}_k = \tilde{\boldsymbol{h}}_k - \boldsymbol{\delta}_k$ in (5.8a) and noticing that $\tilde{\boldsymbol{B}}^H \tilde{\boldsymbol{h}}_k = \boldsymbol{0}$ by construction yields:

$$\lambda_b^{s,k} = \mathrm{E}\left\{ \|\tilde{\boldsymbol{B}}_b^H \boldsymbol{\delta}_k\|^2 \right\}. \tag{5.9}$$

The total blocking ability is defined as the sum of the blocking abilities of all constrained ATFs:

$$\eta_b = \sum_{k=1}^{K} \eta_b^k. \tag{5.10}$$

5.1.3.1 Blocking ability and signal leakage of the eigen-space BM

The noisy eigen-space BM is given by substituting the noisy ATFs (5.6) in (5.2):

$$\tilde{\boldsymbol{B}}_e = \boldsymbol{I}_{M\times M} - \tilde{\boldsymbol{H}}\left(\tilde{\boldsymbol{H}}^H\tilde{\boldsymbol{H}}\right)^{-1}\tilde{\boldsymbol{H}}^H. \tag{5.11}$$

The noise power at the output of the eigen-space BM, λ_e^n, is given by substituting (5.11) in (5.8b):

$$\lambda_e^n = \mathrm{E}\left\{\boldsymbol{w}^H\tilde{\boldsymbol{B}}_e\tilde{\boldsymbol{B}}_e^H\boldsymbol{w}\right\} = \mathrm{trace}\left\{\mathrm{E}\left\{\tilde{\boldsymbol{B}}_e\tilde{\boldsymbol{B}}_e^H\right\}\right\}. \tag{5.12}$$

As $\tilde{\boldsymbol{B}}_e$ is a hermitian projection matrix, the following equation holds: $\tilde{\boldsymbol{B}}_e\tilde{\boldsymbol{B}}_e^H = \tilde{\boldsymbol{B}}_e\tilde{\boldsymbol{B}}_e = \tilde{\boldsymbol{B}}_e$. And after some matrix manipulation λ_e^n equals:

$$\lambda_e^n = M - K. \tag{5.13}$$

The signal leakage of the kth ATF at the output of the eigen-space BM, $\lambda_e^{s,k}$, is given by substituting $\tilde{\boldsymbol{B}}_e$ in (5.9):

$$\lambda_e^{s,k} = \mathrm{E}\left\{\boldsymbol{\delta}_k^H\tilde{\boldsymbol{B}}_e\tilde{\boldsymbol{B}}_e^H\boldsymbol{\delta}_k\right\} \tag{5.14}$$

Expanding (5.14) to a Taylor series around $\boldsymbol{H}$ as a function of $\boldsymbol{\Delta}$, neglecting elements of order $\boldsymbol{\Delta}^n$ for $n > 2$, and using $\mathrm{E}\{\boldsymbol{\Delta}\} = \boldsymbol{0}$, the following approximation holds:

$$\lambda_e^{s,k} \approx \mathrm{E}\left\{\boldsymbol{\delta}_k^H\boldsymbol{B}_e\boldsymbol{B}_e^H\boldsymbol{\delta}_k\right\} = \lambda_u\,\mathrm{trace}\left\{\boldsymbol{B}_e\boldsymbol{B}_e^H\right\}. \tag{5.15}$$

And similarly to the derivation of (5.13), λ_e^s equals:

$$\lambda_e^{s,k} = (M-K)\lambda_u. \tag{5.16}$$

Therefore, the ability of the noisy eigen-space BM $\tilde{\boldsymbol{B}}_e$ to block $\boldsymbol{h}_k$, the ATF of the kth source is given by:

$$\eta_e^k = \frac{\lambda_e^{s,k}}{\lambda_e^n} = \lambda_u. \tag{5.17}$$

And the total blocking ability of the eigen-space BM is:

$$\eta_e = K\lambda_u. \tag{5.18}$$

5.1.3.2 Blocking ability and signal leakage of the sparse BM

The noisy sparse BM is constructed by substituting $\tilde{\boldsymbol{H}}$ (5.6), the noisy estimate of $\boldsymbol{H}$, in (5.5):

$$\tilde{\boldsymbol{B}}_s = \begin{bmatrix} -\tilde{\boldsymbol{\beta}}_{K+1} & \cdots & -\tilde{\boldsymbol{\beta}}_M \\ & \boldsymbol{I}_{(M-K)\times(M-K)} & \end{bmatrix} \tag{5.19}$$

where

$$\tilde{\boldsymbol{\beta}}_m = \left(\tilde{\boldsymbol{H}}_r^{-1}\right)^H \tilde{\boldsymbol{H}}_{m,:}^H \tag{5.20a}$$

$$\tilde{\boldsymbol{H}}_r = \boldsymbol{H}_r + \boldsymbol{\Delta}_r. \tag{5.20b}$$

and $\boldsymbol{\Delta}_r = \Delta_{1:K,1:K}$.

Similarly to the derivation in (5.12), the noise power at the output of the sparse BM is given by:

$$\lambda_s^n = \text{trace}\left\{\text{E}\left\{\tilde{\boldsymbol{B}}_s \tilde{\boldsymbol{B}}_s^H\right\}\right\}. \tag{5.21}$$

Following the definition in (5.19), the latter expression is:

$$\lambda_s^n = M - K + \sum_{m=1}^{M-K} \text{E}\left\{\|\tilde{\boldsymbol{\beta}}_m\|^2\right\}. \tag{5.22}$$

Consider a single term of the sum in (5.22):

$$\text{E}\left\{\|\tilde{\boldsymbol{\beta}}_m\|^2\right\} = \text{E}\left\{\tilde{\boldsymbol{H}}_{m,:}\tilde{\boldsymbol{H}}_r^{-1}\left(\tilde{\boldsymbol{H}}_r^H\tilde{\boldsymbol{H}}_r\right)^{-1}\tilde{\boldsymbol{H}}_{m,:}^H\right\}. \tag{5.23}$$

Assuming again that $\boldsymbol{H}_r$ is invertible and high estimation accuracy, i.e. $\|\boldsymbol{H}_r\|^2 \gg \|\boldsymbol{\Delta}_r\|^2$, and by replacing the expression $\left(\tilde{\boldsymbol{H}}_r^H\tilde{\boldsymbol{H}}_r\right)^{-1}$ with its first term Taylor series expansion around

$\mathbf{0}$, we obtain:

$$\left(\tilde{\boldsymbol{H}}_r^H \tilde{\boldsymbol{H}}_r\right)^{-1} \approx \left(\boldsymbol{I} - \left(\boldsymbol{H}_r^H \boldsymbol{H}_r\right)^{-1} \left(\boldsymbol{H}_r^H \boldsymbol{\Delta}_r + \boldsymbol{\Delta}_r^H \boldsymbol{H}_r + \boldsymbol{\Delta}_r^H \boldsymbol{\Delta}_r\right)\right) \cdot \left(\boldsymbol{H}_r^H \boldsymbol{H}_r\right)^{-1}. \tag{5.24}$$

Next, substituting the approximation (5.24) in (5.23) and neglecting terms $\boldsymbol{\Delta}^n$ of order $n > 2$ and using $\mathrm{E}\{\boldsymbol{\Delta}\} = \mathbf{0}$, the following approximation holds:

$$\begin{aligned}\mathrm{E}\left\{\|\tilde{\boldsymbol{\beta}}_m\|^2\right\} \approx & \|\boldsymbol{\beta}_m\|^2 + \lambda_u \operatorname{trace}\left\{\left(\boldsymbol{H}_r^H \boldsymbol{H}_r\right)^{-1}\right\} \\ & - \lambda_u \boldsymbol{H}_{m,:} \left(\boldsymbol{H}_r^H \boldsymbol{H}_r\right)^{-2} \boldsymbol{H}_{m,:}^H.\end{aligned} \tag{5.25}$$

Finally, substituting (5.25) in (5.22) yields:

$$\begin{aligned}\lambda_s^n =& M - K + \lambda_u \operatorname{trace}\left\{\left(\boldsymbol{H}_r^H \boldsymbol{H}_r\right)^{-1}\right\} (M-K) \\ & + \sum_{m=1}^{M-K} \|\boldsymbol{\beta}_m\|^2 - \lambda_u \boldsymbol{H}_{m+K,:} \left(\boldsymbol{H}_r^H \boldsymbol{H}_r\right)^{-2} \boldsymbol{H}_{m+K,:}^H \\ \approx & M - K + \sum_{m=1}^{M-K} \|\boldsymbol{\beta}_m\|^2\end{aligned} \tag{5.26}$$

where the approximation in the last transition is due to the high estimation accuracy.

Similarly to the derivation of (5.14), the leakage of the kth ATF to the output of the sparse BM is given by:

$$\begin{aligned}\lambda_s^{s,k} =& \lambda_u \operatorname{trace}\left\{\boldsymbol{B}_s \boldsymbol{B}_s^H\right\} \\ =& \left(M - K + \sum_{m=1}^{M-K} \|\boldsymbol{\beta}_m\|^2\right) \lambda_u.\end{aligned} \tag{5.27}$$

The ability of the noisy sparse BM $\tilde{\boldsymbol{B}}_s$ to block $\boldsymbol{h}_k$, the ATF of the kth source is given by:

$$\eta_s^k = \frac{\lambda_s^{s,k}}{\lambda_s^n} = \lambda_u \tag{5.28}$$

and the total blocking ability of the sparse BM is therefore:

$$\eta_s = K \lambda_u. \tag{5.29}$$

Please note that the blocking ability of the proposed sparse BM is equivalent to the blocking ability of the eigen-space BM (5.17), (5.18).

5.1.4 Experimental study

The performance of the proposed sparse BM, and of the eigen-space BM is presented for narrowband signal scenarios in Sec. 5.1.4.1, and for wideband speech signals in Sec. 5.1.4.2.

5.1.4.1 Narrowband signals

A comprehensive Monte-Carlo simulation was performed for validating the theoretical analysis derived in Sec. 5.1.3. A total of 561 scenarios were tested, the parameters of the scenario were: 1) the number of microphones was set to $M = 5, 10, .., 30$; 2) the number of constraints was set to $K = 1, 2, .., \lfloor \frac{M}{2} \rfloor$; 3) the estimation accuracy level was set to $0\text{dB}, 5\text{dB}, .., 50\text{dB}$. At each scenario the performance was averaged over 100 randomly generated ATFs ($\boldsymbol{H}$), times 1000 random estimation errors per instance. Altogether, the blocking abilities of 56.1×10^6 sparse and eigen-space BMs were evaluated and compared with the theoretical analysis. In Fig. 5.1 the average differences between the theoretical blocking ability and of the empirical blocking ability for the sparse and eigen-space BMs are depicted. In these figures the number of microphones was set to $M = 20$, while the numbers of constraints varied in the range of $K = 1, 2, .., \lfloor \frac{M}{2} \rfloor$ and the SNR levels varied in the range of $0\text{dB}, 5\text{dB}, .., 50\text{dB}$. The results validate the theoretical analysis as the average deviation from the theory for estimation accuracies higher than 5dB is lower than 0.5dB.

5.1.4.2 Speech signals

The eigen-space and sparse BMs were tested on wideband speech signals in a simulated $4\text{m} \times 3\text{m} \times 3\text{m}$ room environment with a reverberation time of $T_{60} = 150\text{ms}$. A uniform linear microphone array comprising 9 microphones with 5cm spacing was placed next to one of the walls. Three speakers and a stationary interference were located in the room, at a distance of 1.8m in front of the microphone array, at angles $-60^o, -20^o, 20^o, 60^o$. The received signals were sampled at a sample rate of 8kHz and transformed to the STFT domain with 4096 DFT points and a 50% overlap between frames. The three speakers were constrained. The BMs were calculated in the STFT domain based on the normalized ATFs of the three speakers contaminated by a -30dB error level. High estimation accuracy can be obtained

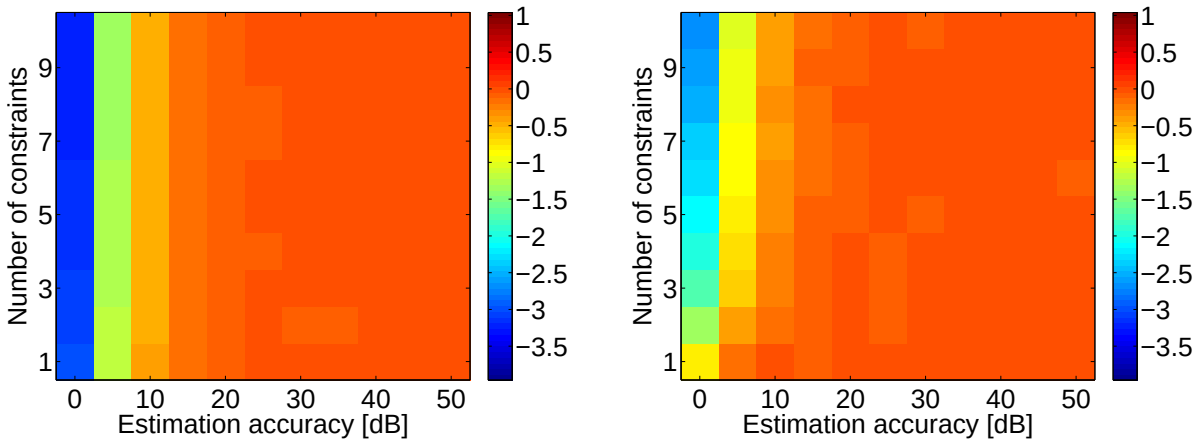


Figure 5.1: Difference in dB between theoretical and empirical blocking abilities for narrowband signals simulation with $M = 20$ microphones, for the eigen-space BM (left) and sparse BM (right).

by applying the subspaces based estimation in [51]. In order to keep the power at the output of the sparse BM at a constant level over frequency, a normalized BM $\tilde{\boldsymbol{B}}_s/\|\tilde{\boldsymbol{B}}_s\|_{\mathrm{F}}$ was used rather than $\tilde{\boldsymbol{B}}_s$. Note that the latter scaling does not affect the blocking ability as the signal leakage and the spatially white noise gain are multiplied by the same factor. The total blocking ability of the eigen-space BM was -28dB while the total blocking ability of sparse BM was slightly worse at -26.5dB. The 1st source as received by the microphone array and its contribution to the leakage at the outputs of the BMs are depicted in Fig. 5.2. Note the different scale in the microphone and leakage figures. It can be verified that the proposed sparse BM, and the eigen-space BM obtain similar performance for wideband speech signals.

5.1.5 Conclusions

A novel systematic scheme for constructing a K constraints sparse BM for the LCMV-BF was derived. The signal leakage and the blocking ability of the proposed sparse BM and of the commonly used eigen-space BM are analyzed and compared. It is analytically proven that the blocking abilities of both BMs are equivalent, provided that the estimation accuracy is high. The computational complexity of the proposed sparse BM is $K \times (M - K)$, which is substantially lower than the computational complexity of the eigen-space BM, which is M^2. The theoretical analysis is experimentally verified for both narrowband signals and wideband

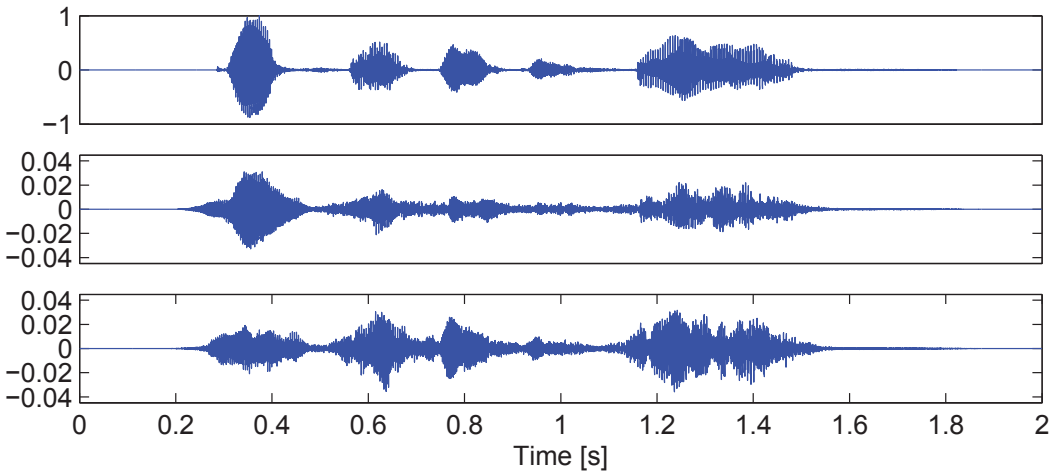


Figure 5.2: Source number 1 as received by the 1st microphone (top), its contribution to the leakage at the output of the eigen-space BM (middle) and at the output of the sparse BM (bottom).

speech signals.

5.2 Low-complexity addition or removal of sensors/constraints in LCMV beamformers

Here, we address the problem of reducing the computational burden of recalculating the LCMV-BF when modifying the group of sensors which participate in the spatial filtering, denoted as the active sensors or nodes, or when modifying the constraints set. We assume that the required BF updates are subjected to a controlling mechanism, referred to as the *oracle*. The decisions of the oracle can be motivated by optimizing the tradeoff between performance and resource usage, handling arbitrary link failures, and also by determining the desired response for the various sources, which result in updating of the constraints set. Updating the configuration of the active sensors could affect the desired constraints set. For example, adding sensors increases the dimension of the received signals and allows for the application of a larger number of constraints. The decision mechanism of the oracle is out of the scope of the current contribution.

In this section, we propose a set of lower complexity procedures for updating the group of active sensors, and the constraints set to a given LCMV-BF. We derive the updating procedures for both the LCMV closed-form BF and its respective GSC form. The proposed

procedures reduce the computational complexity, and are equivalent to the straightforward calculation of the LCMV-BF.

The section is organized as follows. In Sec. 5.2.1 the problem is formulated. In Sec. 5.2.2 four examples for updating procedures of the LCMV-BF are fully derived. Later in the Sec. 5.2.5 all eight updating procedures are summarized. In Sec. 5.2.3 we discuss extending the derived algorithms for adding or removing a group of sensors or constraints. The computational complexity of the proposed procedures is analyzed and compared with the complexity of their corresponding straightforward BFs in Sec. 5.2.4.

5.2.1 Problem formulation

Consider P point source signals, some stationary and other non-stationary, denoted by $s^1(\ell,k),\ldots,s^P(\ell,k)$, propagating in a multi-path environment and impinging on an array comprising M sensors. The problem is formulated using a narrow-band model in the STFT domain, where ℓ is the frame index and k is the frequency index. From hereon, the frequency notation is omitted for brevity. The application and the calculation of the BF should be interpreted frequency-wise. The transfer function (TF) relating the pth source and the mth sensor is denoted by $h_m^p(\ell)$. Define in vector notation:

$$\boldsymbol{s}(\ell) = \begin{bmatrix} s^1(\ell) & \cdots & s^P(\ell) \end{bmatrix}^T \tag{5.30a}$$

$$\boldsymbol{h}^p(\ell) = \begin{bmatrix} h_1^p(\ell) & \cdots & h_M^p(\ell) \end{bmatrix}^T;\ p=1,\ldots,P \tag{5.30b}$$

$$\boldsymbol{H}(\ell) = \begin{bmatrix} \boldsymbol{h}^1(\ell) & \cdots & \boldsymbol{h}^P(\ell) \end{bmatrix}. \tag{5.30c}$$

The received signals vector and its covariance matrix are given by:

$$\boldsymbol{z}(\ell) = \boldsymbol{H}\boldsymbol{s}(\ell) + \boldsymbol{v}(\ell) \tag{5.31a}$$

$$\begin{aligned}\boldsymbol{\Phi} &= \mathrm{E}\left[\boldsymbol{z}(\ell)\boldsymbol{z}^H(\ell)\right] \\ &= \boldsymbol{H}\boldsymbol{\Sigma}\boldsymbol{H}^H + \boldsymbol{\Phi}_{vv}\end{aligned} \tag{5.31b}$$

where $\boldsymbol{v}(\ell)$ denotes the total received interferences vector of the non coherent signals, $\boldsymbol{\Sigma}$ denotes the diagonal covariance matrix of the coherent sources (assuming they are statistically independent), and $\boldsymbol{\Phi}_{vv} = \mathrm{E}\left[\boldsymbol{v}(\ell)\boldsymbol{v}^H(\ell)\right]$ is the covariance matrix of $\boldsymbol{v}(\ell)$.

Consider a general Pth order constraints set:

$$\boldsymbol{C} = \begin{bmatrix} \boldsymbol{c}^1 & \cdots & \boldsymbol{c}^P \end{bmatrix} \tag{5.32a}$$

$$\boldsymbol{g} = \begin{bmatrix} g^1 & \cdots & g^P \end{bmatrix}^T. \tag{5.32b}$$

The optimization criterion of the LCMV-BF is given by:

$$\boldsymbol{w} = \underset{\boldsymbol{C}^H\boldsymbol{w}=\boldsymbol{g}}{\operatorname{argmin}} \ \boldsymbol{w}^H\boldsymbol{\Phi}\boldsymbol{w}. \tag{5.33}$$

The closed-form LCMV-BF is described in Sec. 5.2.1.1, and the efficient GSC implementation is described in Sec. 5.2.1.2.

5.2.1.1 Closed-form LCMV-BF

This BF form is obtained by solving (5.33) directly using Lagrange multipliers. The closed-form LCMV-BF solution to the problem is given by:

$$\boldsymbol{w} = \boldsymbol{\Phi}^{-1}\boldsymbol{C}\boldsymbol{Q}^{-1}\boldsymbol{g} \tag{5.34}$$

where $\boldsymbol{Q}$ is defined as follows:

$$\boldsymbol{Q} = \boldsymbol{C}^H\boldsymbol{\Phi}^{-1}\boldsymbol{C}. \tag{5.35}$$

We denote the solution in (5.34) as the straightforward LCMV (SF-LCMV). Its computational complexity is mainly dominated by the two matrix inversion $\boldsymbol{\Phi}^{-1}$, and $\boldsymbol{Q}^{-1}$.

In the following sections, we derive algorithms for updating an existing LCMV-BF. We consider two types of updates. The first type is sensor updates and the second type is constraint set updates. For each type of update we derive two procedures. The first procedure, denoted as the incremental procedure, refers to adding either a sensor or a constraint to an existing BF. The second procedure, denoted as the decremental procedure, refers to removing either a sensor or constraint from an existing BF. The derived procedures reduces the dimensions of the matrices to be inverted, and hence reduce the computational complexity substantially.

5.2.1.2 GSC-form LCMV-BF

This BF form is obtained by splitting the applied filters into two components, i.e. $\boldsymbol{w} = \boldsymbol{w}_{\parallel} - \boldsymbol{w}_{\perp}$. The components, $\boldsymbol{w}_{\parallel}$ and $\boldsymbol{w}_{\perp}$, lie in the column-subspace of the constraint matrix $\boldsymbol{C}$ and its complement null-subspace, respectively. The GSC formulation of the problem is given by:

$$\boldsymbol{w} = \boldsymbol{w}_{\parallel} - \boldsymbol{B}\boldsymbol{f} \tag{5.36a}$$

$$\boldsymbol{w}_{\parallel} = \boldsymbol{C}\boldsymbol{R}^{-1}\boldsymbol{g} \tag{5.36b}$$

$$\boldsymbol{f} = \left(\boldsymbol{B}^{H}\boldsymbol{\Phi}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^{H}\boldsymbol{\Phi}\boldsymbol{w}_{\parallel} \tag{5.36c}$$

$$\boldsymbol{R} = \boldsymbol{C}^{H}\boldsymbol{C}. \tag{5.36d}$$

The GSC form is decomposed into two branches. The upper branch, also known as the quiescent BF, is denoted by $\boldsymbol{w}_{\parallel}$. It is responsible for maintaining the constraints set. The lower branch is comprised of two parts: the BM and the subsequent NC denoted by $\boldsymbol{B}$ and $\boldsymbol{f}$, respectively. The objective of the BM is to block the signals arriving from the constraints set subspace and generate $M - P$ interference-only reference signals. Its dimensions are $M \times (M - P)$, and it can be calculated, for example, by applying the SVD to the constraints matrix $\boldsymbol{C}$ [25]. We will assume that all the columns of the BM are orthogonal, as in [51]. Note that an orthogonal BM can always be constructed. The NC uses the reference signals from the output of the BM to estimate the noise component at the output of the quiescent BF and therefore reduce its level. We denote the GSC-form BF in (5.36) as the straightforward GSC (SF-GSC) BF.

The computational complexity of the SF-GSC BF is mainly dominated by the SVD used for constructing the BM, and by the matrix inversion $\boldsymbol{R}^{-1}$. In the following sections, we derive algorithms for updating an existing GSC-BF. We consider two types of updates. The first type is sensor updates and the second type is constraint set updates. For each type of update we derive incremental and decremental procedures which circumvent the SVD and the matrix inversion, and hence reduce the computational complexity substantially. The NC is usually implemented as an ANC using the LMS algorithm [39]. The LMS algorithm consumes $\mathcal{O}(M)$ operations per frequency bin. Due to its low complexity and adaptive nature, it is unnecessary to formulate an update procedure to the ANC.

Please note that in the following sections some notations may be re-defined for brevity.

Explicitly, when considering sensor addition or removal, a constraints set comprising P constraints is assumed. Also, when considering constraint addition or removal, an array comprising M sensors is assumed.

5.2.2 Low-Complexity Beamformer Updating Methods

Algorithms for adding or removing a single constraint to the LCMV-BF and the associated GSC implementation are now derived. The algorithms are denoted by [S\C]U[I\D]-[GSC\LCMV], where S\C stands for sensor or constraint, respectively, U stands for update, I\D stands for incremental or decremental, respectively, and GSC\LCMV stands for the GSC or the closed-form implementations, respectively. For example, the sensor update incremental closed-form implementation algorithm is denoted by SUI-LCMV. For brevity we do not derive all eight algorithms in details. Instead, we chose to elaborate on the derivation of four representative procedures, namely the SUI-LCMV, SUI-GSC, CUI-GSC and the SUD-GSC in the following sub-sections. The derivation of the other algorithms is based on the same methods. A summary of all eight algorithms is given in Sec. 5.2.5.

5.2.2.1 Derivation of the SUI-LCMV Algorithm

Assume an $M-1$ sensors and P constraints LCMV-BF is active.

$$\boldsymbol{w} = \boldsymbol{\Phi}^{-1}\boldsymbol{C}\boldsymbol{Q}^{-1}\boldsymbol{g} \tag{5.37}$$

where $\boldsymbol{C}$ is the $(M-1)\times P$ constraints matrix, $\boldsymbol{g}$ is the $P\times 1$ desired response vector, and the constraints set is

$$\boldsymbol{C}^H\boldsymbol{w} = \boldsymbol{g}. \tag{5.38}$$

Now, a new sensor (indexed M) becomes available. Define the augmented constraints set:

$$\dot{\boldsymbol{C}} = \begin{bmatrix} \boldsymbol{C} \\ \dot{\boldsymbol{c}}^H \end{bmatrix} \tag{5.39}$$

where $\dot{\boldsymbol{c}}$ is a $P\times 1$ vector extending the constraints set to M sensors.

The covariance matrix of the M sensors is given by

$$\dot{\boldsymbol{\Phi}} = \left[\begin{array}{c|c} \boldsymbol{\Phi} & \dot{\boldsymbol{\phi}} \\ \hline \dot{\boldsymbol{\phi}}^H & \dot{\sigma}^2 \end{array}\right]. \tag{5.40}$$

Applying the block matrix inversion formula [90], the inverse of the covariance matrix equals

$$\dot{\boldsymbol{\Phi}}^{-1} = \left[\begin{array}{c|c} \boldsymbol{\Phi}^{-1} + \dot{\epsilon}\boldsymbol{\Phi}^{-1}\dot{\boldsymbol{\phi}}\dot{\boldsymbol{\phi}}^H\boldsymbol{\Phi}^{-1} & -\dot{\epsilon}\boldsymbol{\Phi}^{-1}\dot{\boldsymbol{\phi}} \\ \hline -\dot{\epsilon}\dot{\boldsymbol{\phi}}^H\boldsymbol{\Phi}^{-1} & \dot{\epsilon} \end{array}\right] \tag{5.41}$$

where

$$\dot{\boldsymbol{\phi}} = \mathrm{E}\left[\left[\begin{array}{ccc} z_1 & \cdots & z_{M-1} \end{array}\right]^T z_M^*\right] \tag{5.42a}$$

$$\dot{\sigma}^2 = \mathrm{E}\left[z_M z_M^*\right] \tag{5.42b}$$

$$\dot{\epsilon} = \left(\dot{\sigma}^2 - \dot{\boldsymbol{\phi}}^H\boldsymbol{\Phi}^{-1}\dot{\boldsymbol{\phi}}\right)^{-1}. \tag{5.42c}$$

Considering the definition of $\boldsymbol{Q}$ in (5.35), the updated $\dot{\boldsymbol{Q}}$ in terms of $\boldsymbol{Q}$ is given by:

$$\begin{aligned} \dot{\boldsymbol{Q}} =& \dot{\boldsymbol{C}}^H\dot{\boldsymbol{\Phi}}^{-1}\dot{\boldsymbol{C}} \\ =& \left[\begin{array}{c|c} \boldsymbol{C}^H & \dot{\boldsymbol{c}} \end{array}\right] \\ & \cdot \left[\begin{array}{c|c} \boldsymbol{\Phi}^{-1} + \dot{\epsilon}\boldsymbol{\Phi}^{-1}\dot{\boldsymbol{\phi}}\dot{\boldsymbol{\phi}}^H\boldsymbol{\Phi}^{-1} & -\dot{\epsilon}\boldsymbol{\Phi}^{-1}\dot{\boldsymbol{\phi}} \\ \hline -\dot{\epsilon}\dot{\boldsymbol{\phi}}^H\boldsymbol{\Phi}^{-1} & \dot{\epsilon} \end{array}\right] \\ & \cdot \left[\begin{array}{c} \boldsymbol{C} \\ \hline \dot{\boldsymbol{c}}^H \end{array}\right] \qquad (5.43) \\ =& \boldsymbol{Q} + \dot{\epsilon}\dot{\boldsymbol{q}}\dot{\boldsymbol{q}}^H \qquad (5.44) \end{aligned}$$

where

$$\dot{\boldsymbol{q}} = \boldsymbol{C}^H\boldsymbol{\Phi}^{-1}\dot{\boldsymbol{\phi}} - \dot{\boldsymbol{c}}. \tag{5.45}$$

Applying the Woodbury identity [119] to the inverse of (5.44), $\dot{\boldsymbol{Q}}^{-1}$ equals

$$\dot{\boldsymbol{Q}}^{-1} = \boldsymbol{Q}^{-1} - \frac{\boldsymbol{Q}^{-1}\dot{\boldsymbol{q}}\dot{\boldsymbol{q}}^H\boldsymbol{Q}^{-1}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H\boldsymbol{Q}^{-1}\dot{\boldsymbol{q}}}. \tag{5.46}$$

Finally, the updated BF, $\dot{\boldsymbol{w}}$, is given in terms of the previous BF terms, $\boldsymbol{w}$, $\boldsymbol{Q}$, $\boldsymbol{Q}^{-1}$, $\boldsymbol{\Phi}$ and

$\boldsymbol{\Phi}^{-1}$, by substituting (5.41), (5.46) in (5.34):

$$\dot{\boldsymbol{w}} = \begin{bmatrix} \boldsymbol{w} + \dot{\boldsymbol{\Delta w}} \\ \dot{w}_M \end{bmatrix} \tag{5.47a}$$

$$\dot{\boldsymbol{\Delta w}} = -\frac{\dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \boldsymbol{g}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}}} \boldsymbol{\Phi}^{-1} \boldsymbol{C} \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}} - \dot{w}_M \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}} \tag{5.47b}$$

$$\begin{aligned} \dot{w}_M = &- \dot{\epsilon} \left(\dot{\boldsymbol{\phi}}^H \boldsymbol{\Phi}^{-1} \boldsymbol{C} \boldsymbol{Q}^{-1} - \dot{\boldsymbol{c}}^H \boldsymbol{Q}^{-1} \right) \\ &\cdot \left(\boldsymbol{I}_{P \times P} - \frac{\dot{\boldsymbol{q}} \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}}} \right) \boldsymbol{g}. \end{aligned} \tag{5.47c}$$

A block-diagram of the SUI-LCMV algorithm is depicted in Fig. 5.3. The procedure is summarized in Alg. 2.

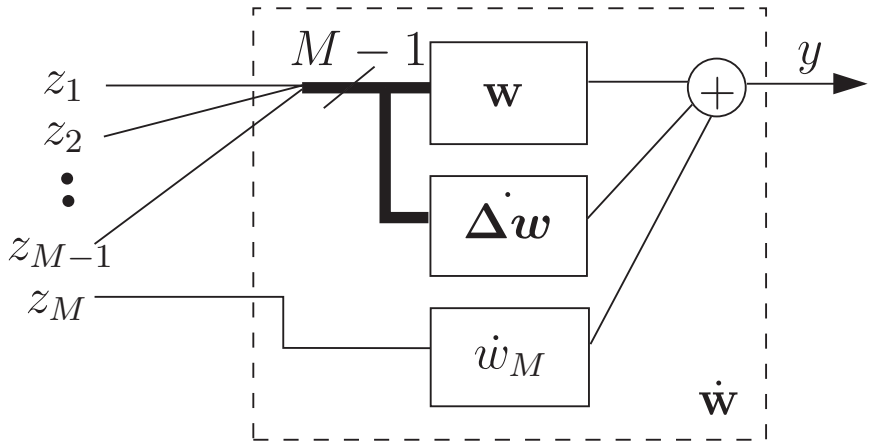


Figure 5.3: Block-diagram of the SUI-LCMV procedure.

5.2.2.2 Derivation of the SUI-GSC Algorithm

Similarly to (5.36), suppose now that an $M-1$ sensors GSC BF maintaining P constraints is given by:

$$\boldsymbol{w} = \boldsymbol{w}_{\parallel} - \boldsymbol{B}\boldsymbol{f} \tag{5.48a}$$

$$\boldsymbol{w}_{\parallel} = \boldsymbol{C}\boldsymbol{R}^{-1}\boldsymbol{g} \tag{5.48b}$$

$$\boldsymbol{f} = \left(\boldsymbol{B}^H \boldsymbol{\Phi} \boldsymbol{B}\right)^{-1} \boldsymbol{B}^H \boldsymbol{\Phi} \boldsymbol{w}_{\parallel} \tag{5.48c}$$

$$\boldsymbol{R} = \boldsymbol{C}^H \boldsymbol{C} \tag{5.48d}$$

where we assume that the ANC has converged to $\boldsymbol{f}$. The latter filter is the appropriate Wiener filter for estimating the noise component at the output of the quiescent BF based on the noise references at the output of the BM. We further assume that $\boldsymbol{B}$ is an appropriate $(M-1)\times(M-1-P)$ BM. The BM can be calculated for example using the SVD of $\boldsymbol{C}$ [25]. We assume that the BM is orthogonal, i.e. $\boldsymbol{B}^H\boldsymbol{B} = \boldsymbol{I}_{(M-1-P)\times(M-1-P)}$.

Consider adding the Mth sensor and updating the BF. The updated constraints set is defined as in (5.39). The updated $\dot{\boldsymbol{R}}$ matrix is given by substituting (5.39) in (5.48d)

$$\dot{\boldsymbol{R}} = \boldsymbol{R} + \dot{\boldsymbol{c}}\dot{\boldsymbol{c}}^H. \tag{5.49}$$

Applying the Woodbury identity to the inverse of (5.49), $\dot{\boldsymbol{R}}^{-1}$ is given by

$$\dot{\boldsymbol{R}}^{-1} = \boldsymbol{R}^{-1} - \frac{\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}\dot{\boldsymbol{c}}^H\dot{\boldsymbol{R}}^{-1}}{1+\dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}}. \tag{5.50}$$

The M sensors quiescent BF is given, similarly to (5.48b), by replacing $\boldsymbol{C}$ and $\boldsymbol{R}^{-1}$ with $\dot{\boldsymbol{C}}$ and $\dot{\boldsymbol{R}}^{-1}$ from equations (5.39), (5.50):

$$\dot{\boldsymbol{w}}_{\parallel} = \begin{bmatrix} \boldsymbol{w}_{\parallel} + \dot{\boldsymbol{\Delta}}\boldsymbol{w}_{\parallel} \\ \dot{w}_{\parallel M} \end{bmatrix} \tag{5.51a}$$

$$\dot{w}_{\parallel M} = \frac{\dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\boldsymbol{g}}{1+\dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}} \tag{5.51b}$$

$$\dot{\boldsymbol{\Delta}}\boldsymbol{w}_{\parallel} = -\dot{w}_{\parallel M}\boldsymbol{C}\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}. \tag{5.51c}$$

Next, we address the problem of updating the BM. Since we added the Mth sensor, there should be $M-P$ signals at the output of the BM. The updated BM, $\dot{\boldsymbol{B}}$, should block the signal subspace, i.e. $\dot{\boldsymbol{B}}^H\dot{\boldsymbol{C}} = \boldsymbol{0}$. The first $M-P-1$ reference signals are equivalent to the older ones. This can be verified by adding a row of zeros to $\boldsymbol{B}$, i.e. $\begin{bmatrix} \boldsymbol{B} \\ \boldsymbol{0}_{1\times(M-P-1)} \end{bmatrix}^H \dot{\boldsymbol{C}} = \boldsymbol{B}^H\boldsymbol{C} = \boldsymbol{0}_{(M-P-1)\times P}$. We suggest to use

$$\dot{\boldsymbol{\Delta}}\boldsymbol{b} = \frac{\left[\begin{array}{c|c} \dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\boldsymbol{C}^H & -1 \end{array}\right]^H}{\| \left[\begin{array}{c|c} \dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\boldsymbol{C}^H & -1 \end{array}\right] \|} \tag{5.52}$$

as the $(M-P)$th column of the updated BM. $\dot{\boldsymbol{\Delta}}\boldsymbol{b}$ is orthogonal to the first $M-P-1$

columns of $\dot{\boldsymbol{B}}$ since

$$\begin{aligned}\begin{bmatrix} \boldsymbol{B} \\ \mathbf{0}_{1\times(M-P-1)} \end{bmatrix}^H \dot{\boldsymbol{\Delta b}} =& \frac{\boldsymbol{B}^H \boldsymbol{C} \boldsymbol{R}^{-1} \dot{\boldsymbol{c}}}{\| \begin{bmatrix} \dot{\boldsymbol{c}}^H \boldsymbol{R}^{-1} \boldsymbol{C}^H & -1 \end{bmatrix} \|} - \mathbf{0}_{(M-P-1)\times 1} \\ =& \mathbf{0}_{(M-P-1)\times 1}\end{aligned}$$

where the last transition is again due to $\boldsymbol{B}^H \boldsymbol{C} = \mathbf{0}_{(M-P-1)\times P}$. $\dot{\boldsymbol{\Delta b}}$ is also orthogonal to $\dot{\boldsymbol{C}}$ since:

$$\begin{aligned}\dot{\boldsymbol{\Delta b}}^H \dot{\boldsymbol{C}} =& \frac{\begin{bmatrix} \dot{\boldsymbol{c}}^H \boldsymbol{R}^{-1} \boldsymbol{C}^H & -1 \end{bmatrix} \begin{bmatrix} \boldsymbol{C} \\ \dot{\boldsymbol{c}}^H \end{bmatrix}}{\| \begin{bmatrix} \dot{\boldsymbol{c}}^H \boldsymbol{R}^{-1} \boldsymbol{C}^H & -1 \end{bmatrix} \|} \\ =& \frac{\dot{\boldsymbol{c}}^H \boldsymbol{R}^{-1} \boldsymbol{C}^H \boldsymbol{C} - \dot{\boldsymbol{c}}^H}{\| \begin{bmatrix} \dot{\boldsymbol{c}}^H \boldsymbol{R}^{-1} \boldsymbol{C}^H & -1 \end{bmatrix} \|} \\ =& \mathbf{0}_{1\times P}\end{aligned}$$

where the last transition is due to the definition of $\boldsymbol{R}$ in (5.48d). Therefore, augmenting $\boldsymbol{B}$ by $\dot{\boldsymbol{\Delta b}}$ is a proper BM of P constraints:

$$\dot{\boldsymbol{B}} = \left[\begin{array}{c|c} \boldsymbol{B} & \dot{\boldsymbol{\Delta b}} \\ \mathbf{0}_{1\times M-1-P} & \end{array} \right]. \tag{5.53}$$

After updating the quiescent BF and the BM, another reference signal is added. In the general case the new reference signal and the previous reference signals are correlated. Therefore, not only the NC filter of the new reference signal needs to be determined, but also the NC filters of the previous reference signals need to be adjusted. As mentioned earlier, we rely on the low complexity and fast convergence of the LMS algorithm for updating the NC coefficients. The resulting NC after convergence is given by substituting (5.40), (5.51a), (5.53) in (5.48c):

$$\dot{\boldsymbol{f}} = \left(\dot{\boldsymbol{B}}^H \dot{\boldsymbol{\Phi}} \dot{\boldsymbol{B}} \right)^{-1} \dot{\boldsymbol{B}}^H \dot{\boldsymbol{\Phi}} \dot{\boldsymbol{w}}_{\|}. \tag{5.54}$$

A block-diagram of the SUI-GSC algorithm is depicted in Fig. 5.4. The algorithm is summarized in Alg. 6.

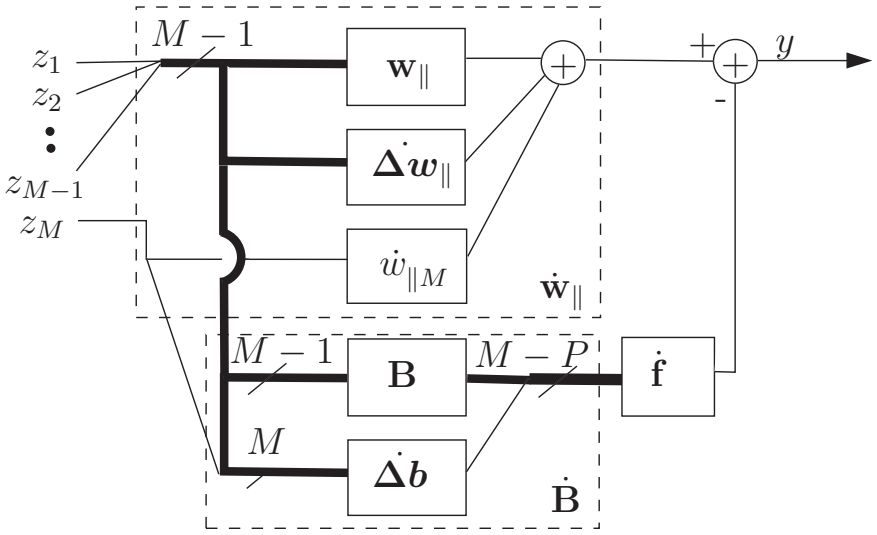


Figure 5.4: Block-Diagram of the SUI-GSC procedure.

5.2.2.3 Derivation of the CUI-GSC Algorithm

Suppose that an M sensors $P-1$ constraints GSC BF is given by:

$$\boldsymbol{w} = \boldsymbol{w}_{\|} - \boldsymbol{B}\boldsymbol{f} \tag{5.55}$$

$$\boldsymbol{w}_{\|} = \boldsymbol{C}\boldsymbol{R}^{-1}\boldsymbol{g} \tag{5.56}$$

$$\boldsymbol{f} = \left(\boldsymbol{B}^H\boldsymbol{\Phi}\boldsymbol{B}\right)^{-1}\boldsymbol{B}^H\boldsymbol{\Phi}\boldsymbol{w}_{\|} \tag{5.57}$$

$$\boldsymbol{R} = \boldsymbol{C}^H\boldsymbol{C} \tag{5.58}$$

where $\boldsymbol{C}$ is the $M \times (P-1)$ constraints matrix, $\boldsymbol{g}$ is the $(P-1)\times 1$ desired response vector and $\boldsymbol{B}$ is an appropriate $M \times (M-P+1)$ BM. As was previously stated, we assume that the BM is orthogonal, i.e. $\boldsymbol{B}^H\boldsymbol{B} = \boldsymbol{I}_{(M-P+1)\times(M-P+1)}$.

Consider adding the Pth constraint and updating the BF. The updated constraints set is

$$\ddot{\boldsymbol{C}} = \left[\begin{array}{c|c} \boldsymbol{C} & \ddot{\boldsymbol{c}} \end{array}\right] \tag{5.59a}$$

$$\ddot{\boldsymbol{g}} = \left[\begin{array}{c|c} \boldsymbol{g}^H & \ddot{g}^* \end{array}\right]^H. \tag{5.59b}$$

Updating the matrix $\boldsymbol{R}$ in (5.58) with the Pth constraint yields

$$\ddot{\boldsymbol{R}} = \left[\begin{array}{c|c} \boldsymbol{R} & \ddot{\boldsymbol{r}} \\ \hline \ddot{\boldsymbol{r}}^H & \|\ddot{\boldsymbol{c}}\|^2 \end{array}\right] \tag{5.60}$$

where $\ddot{\boldsymbol{r}} = \boldsymbol{C}^H \ddot{\boldsymbol{c}}$. The inverse of $\ddot{\boldsymbol{R}}$ is given by applying the block matrix inversion formula

$$\ddot{\boldsymbol{R}}^{-1} = \left[\begin{array}{c|c} \boldsymbol{R}^{-1} + \ddot{\rho}\boldsymbol{R}^{-1}\ddot{\boldsymbol{r}}\ddot{\boldsymbol{r}}^H\boldsymbol{R}^{-1} & -\ddot{\rho}\boldsymbol{R}^{-1}\ddot{\boldsymbol{r}} \\ \hline -\ddot{\rho}\ddot{\boldsymbol{r}}^H\boldsymbol{R}^{-1} & \ddot{\rho} \end{array}\right] \tag{5.61}$$

where

$$\ddot{\rho} = \left(\|\ddot{\boldsymbol{c}}\|^2 - \ddot{\boldsymbol{r}}^H\boldsymbol{R}^{-1}\ddot{\boldsymbol{r}}\right)^{-1}. \tag{5.62}$$

The updated quiescent BF designed to maintain the P constraints set is given by substituting the updated values of $\ddot{\boldsymbol{R}}^{-1}$, $\ddot{\boldsymbol{C}}$ and $\ddot{\boldsymbol{g}}$ from (5.61), (5.59a), (5.59b) in (5.56):

$$\ddot{\boldsymbol{w}}_{\|} = \boldsymbol{w}_{\|} + \ddot{\boldsymbol{\Delta}}\boldsymbol{w}_{\|} \tag{5.63a}$$

$$\ddot{\boldsymbol{\Delta}}\boldsymbol{w}_{\|} = \ddot{\rho}\left(\ddot{g} - \ddot{\boldsymbol{c}}^H\boldsymbol{w}_{\|}\right)\left(\boldsymbol{I} - \boldsymbol{C}\boldsymbol{R}^{-1}\boldsymbol{C}^H\right)\ddot{\boldsymbol{c}}. \tag{5.63b}$$

Next, we update the BM. Notice that the rank of the BM equals the number of sensors minus the number of constraints (assuming the constraints set are linearly independent), i.e. $M - P + 1$. Therefore, the rank of the BM corresponding to the modified constraints set is smaller by one than that of the former BM. Hence, we would like to reduce the dimensions of the current BM to $M \times (M - P)$ such that its columns are an orthogonal set and that

$$\ddot{\boldsymbol{B}}^H\ddot{\boldsymbol{C}} = \boldsymbol{0}_{(M-P)\times P}. \tag{5.64}$$

The new constraint vector $\ddot{\boldsymbol{c}}$ can be written as a combination of two components:

$$\ddot{\boldsymbol{c}} = \left(\boldsymbol{I} - \boldsymbol{B}\boldsymbol{B}^H\right)\ddot{\boldsymbol{c}} + \boldsymbol{B}\boldsymbol{B}^H\ddot{\boldsymbol{c}}. \tag{5.65}$$

The first component lies in the $P-1$ constraints subspace, and the second component lies in its corresponding null-subspace, hence spanned by the columns of $\boldsymbol{B}$. The new BM, $\ddot{\boldsymbol{B}}$, should block both $\boldsymbol{C}$ and $\boldsymbol{B}\boldsymbol{B}^H\ddot{\boldsymbol{c}}$, the component of $\ddot{\boldsymbol{c}}$ not spanned by the columns of $\boldsymbol{C}$. This can be obtained by: 1) rotating the current BM such that all but one of its columns

are orthogonal to the second component of $\ddot{\boldsymbol{c}}$; 2) deleting that column. The Householder transformation [120] can be applied to satisfy both requirements. The transformed BM is given by:

$$\tilde{\boldsymbol{B}} = \boldsymbol{B}\left(\boldsymbol{I}_{(M-P+1)\times(M-P+1)} - \frac{2\ddot{\boldsymbol{\chi}}\ddot{\boldsymbol{\chi}}^H}{\|\ddot{\boldsymbol{\chi}}\|^2}\right) \tag{5.66}$$

where $\ddot{\boldsymbol{\chi}}$ is defined as

$$\ddot{\boldsymbol{\chi}} = \frac{\ddot{\boldsymbol{b}}}{\|\ddot{\boldsymbol{b}}\|} + \exp\left(j\angle\ddot{b}_{M-P+1}\right)\boldsymbol{i}_{M-P+1}, \tag{5.67}$$

$\angle(\cdot)$ denotes the angle extraction of a complex number, $\boldsymbol{i}_{M-P+1} = \left[\begin{array}{c|c} \boldsymbol{0}_{1\times(M-P+1)} & 1 \end{array}\right]^T$, $\ddot{\boldsymbol{b}}$ is the projection of $\ddot{\boldsymbol{c}}$ onto $\boldsymbol{B}$, i.e. $\ddot{\boldsymbol{b}} = \boldsymbol{B}^H\ddot{\boldsymbol{c}}$, and $\ddot{b}_{M-P+1}$ is the last entry of $\ddot{\boldsymbol{b}}$. It follows that:

$$\tilde{\boldsymbol{B}}^H\boldsymbol{C} = \boldsymbol{0}_{(M-P+1)\times(P-1)} \tag{5.68a}$$

$$\tilde{\boldsymbol{B}}^H\ddot{\boldsymbol{c}} = -\exp\left(j\angle\ddot{b}_1\right)\boldsymbol{i}_{M-P+1}. \tag{5.68b}$$

Note that since the Householder transformation is unitary, the rotated basis remains orthogonal. The orthogonality property of $\boldsymbol{B}$ is imperative for assuring that all columns of $\tilde{\boldsymbol{B}}$ but the last one are orthogonal to $\boldsymbol{B}\boldsymbol{B}^H\ddot{\boldsymbol{c}}$. Finally, the updated BM is obtained by deleting the last column of $\tilde{\boldsymbol{B}}$:

$$\ddot{\boldsymbol{B}} = \tilde{\boldsymbol{B}}\bar{\boldsymbol{I}}_{M-P}, \tag{5.69}$$

where $\bar{\boldsymbol{I}}_m$ is an $(m-1)\times m$ matrix constructed by removing the last row of the identity matrix $\boldsymbol{I}_{m\times m}$.

In a similar manner to the NC update of the SUI-GSC procedure in Sec. 5.2.2.2, the updated NC filters after convergence are given in a vector form by substituting (5.69), (5.31b), (5.63a) in (5.57):

$$\ddot{\boldsymbol{f}} = \left(\ddot{\boldsymbol{B}}^H\boldsymbol{\Phi}\ddot{\boldsymbol{B}}\right)^{-1}\ddot{\boldsymbol{B}}^H\boldsymbol{\Phi}\ddot{\boldsymbol{w}}_{\|}. \tag{5.70}$$

A block-diagram of the CUI-GSC algorithm is depicted in Fig. 5.5. The algorithm is summarized in Alg. 8.

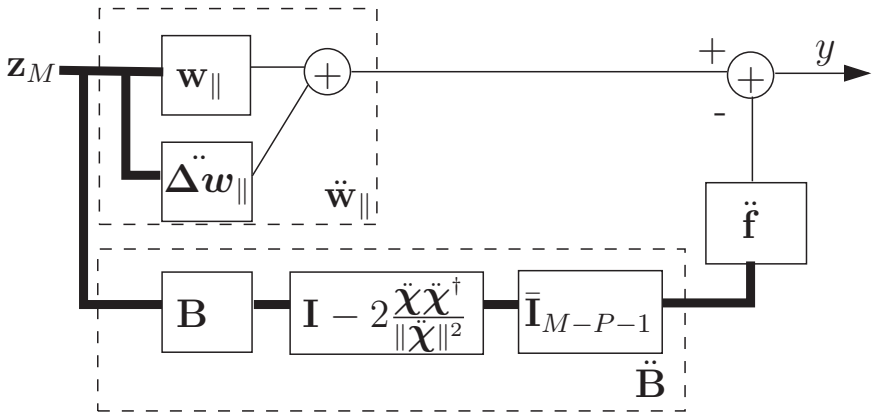


Figure 5.5: Block-diagram of the CUI-GSC procedure.

5.2.2.4 Derivation of the SUD-GSC Algorithm

Suppose that an M sensors and P constraints GSC-BF is given by:

$$\dot{\boldsymbol{w}} = \dot{\boldsymbol{w}}_{\|} - \dot{\boldsymbol{B}}\dot{\boldsymbol{f}} \tag{5.71a}$$

$$\dot{\boldsymbol{w}}_{\|} = \dot{\boldsymbol{C}}\dot{\boldsymbol{R}}^{-1}\boldsymbol{g} \tag{5.71b}$$

where $\dot{\boldsymbol{C}}$, $\dot{\boldsymbol{R}}$ are defined as in (5.39), (5.49), respectively, and $\dot{\boldsymbol{B}}$ is an $M \times (M-P)$ orthogonal BM. Now, consider that the Mth sensor becomes unavailable. In this sub-section we derive the equations for updating the BF using its previous value. The updated $\boldsymbol{R}^{-1}$ is given by applying the Woodbury identity to the inverse of (5.49):

$$\boldsymbol{R}^{-1} = \dot{\boldsymbol{R}}^{-1} + \frac{\dot{\boldsymbol{R}}^{-1}\dot{\boldsymbol{c}}\dot{\boldsymbol{c}}^{H}\dot{\boldsymbol{R}}^{-1}}{1 - \dot{\boldsymbol{c}}^{H}\dot{\boldsymbol{R}}^{-1}\dot{\boldsymbol{c}}}. \tag{5.72}$$

Substituting (5.72), (5.39) in (5.71b) yields:

$$\boldsymbol{w}_{\|} = \bar{\boldsymbol{I}}_M^H \dot{\boldsymbol{w}}_{\|} - \dot{\boldsymbol{\Delta}}\boldsymbol{w}_{\|} \tag{5.73a}$$

$$\dot{\boldsymbol{\Delta}}\boldsymbol{w}_{\|} = -\dot{w}_{\|M}\boldsymbol{C}\boldsymbol{R}^{-1}\dot{\boldsymbol{c}} \tag{5.73b}$$

$$\dot{w}_{\|M} = \dot{\boldsymbol{i}}_M^H \dot{\boldsymbol{w}}_{\|}. \tag{5.73c}$$

Next we address updating the BM. We apply the Householder transformation step and

diagonalize the last row of $\dot{\boldsymbol{B}}$. Define:

$$\dot{\boldsymbol{\chi}} = \dot{\boldsymbol{B}}^H \boldsymbol{i}_M + \|\dot{\boldsymbol{B}}^H \boldsymbol{i}_M\| \exp\left(-j\angle \dot{B}_{M,M-P}\right) \boldsymbol{i}_{M-P} \tag{5.74}$$

where $\dot{B}_{M,M-P}$ is the $(M, M-P)$ entry in $\dot{\boldsymbol{B}}$. The rotated BM is given by:

$$\tilde{\boldsymbol{B}} = \dot{\boldsymbol{B}} \left(\boldsymbol{I} - \frac{2\dot{\boldsymbol{\chi}}\dot{\boldsymbol{\chi}}^H}{\|\dot{\boldsymbol{\chi}}\|^2} \right). \tag{5.75}$$

It can be verified that the last row of $\tilde{\boldsymbol{B}}$ equals:

$$\boldsymbol{i}_M^H \tilde{\boldsymbol{B}} = \left[\; \boldsymbol{0}_{1\times(M-P-1)} \;\middle|\; -\|\dot{\boldsymbol{B}}^H \boldsymbol{i}_M\| \exp\left(j\angle \dot{B}_{M,M-P}\right) \; \right]. \tag{5.76}$$

Since the Householder transformation is unitary, the orthogonality of the BM is kept. The rotated matrix keeps blocking the original constraints matrix, i.e. $\tilde{\boldsymbol{B}}^H \dot{\boldsymbol{C}} = \boldsymbol{0}_{(M-P)\times P}$. Finally, the updated $(M-1)\times(M-P-1)$ dimensional BM is obtained by deleting the last row and last column of $\tilde{\boldsymbol{B}}$

$$\boldsymbol{B} = \bar{\boldsymbol{I}}_M^H \tilde{\boldsymbol{B}} \bar{\boldsymbol{I}}_{M-P} \tag{5.77}$$

and the NC is given after convergence by (5.36c).

A block-diagram of the SUD-GSC algorithm is depicted in Fig. 5.6. The algorithm is summarized in Alg. 7.

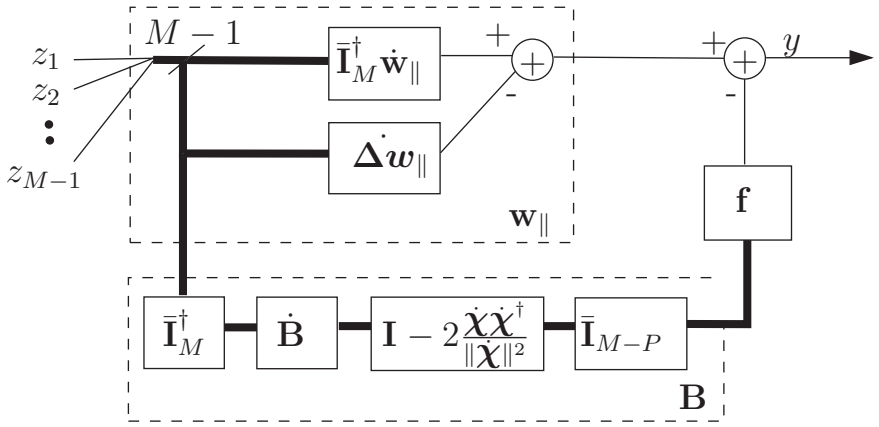


Figure 5.6: Block-diagram of the SUD-GSC procedure.

5.2.3 Adding/removing a group of sensors/constraints

In a partially connected sensor network some nodes are accessible only indirectly through some other nodes. In these networks a change in a single link can affect the activity of multiple nodes. Here we discuss adding or removing multiple nodes or multiple constraints to an LCMV-BF.

Two basic methods are used by the algorithms derived in this section. The first method is the block matrix inversion formula [90] which is used for inverting an $m \times m$ matrix based on the already calculated inverse of an $(m-1) \times (m-1)$ sub-matrix. Two strategies can be adopted in the application of the block matrix inversion formula in cases of $k > 1$ sensor/constraint updates. One strategy utilizes k sequential updates as derived previously. An alternative strategy uses the more general version of the block matrix inversion formula. Namely, the inverse of the $(m-k) \times (m-k)$ sub-matrix is utilized in the inversion of the $m \times m$ matrix. The latter strategy results in more cumbersome expressions. As both strategies involve equivalent computational burden, the sequential strategy of multiple updates is preferred.

The second method used in this section, is the Householder transformation step [120], which is used for rotating an orthogonal basis such that all of its new basis vectors but one are orthogonal to a predefined vector. A sequence of Householder transformation steps can be applied for multiple sensor/constraint updates. The detailed derivation of these algorithms, as well as their complexity analysis, is out of the scope of the current contribution.

5.2.4 Complexity evaluation

In this section we compare the complexity of the straightforward LCMV-BF closed-form and GSC form implementations with their updated form counterparts. Opposed to the straightforward BFs, the updating procedures rely on calculation results of previous BFs, and therefore impose memory requirements. We consider both computational complexity and memory requirements. The computational analysis is based on the complexity of basic operations [121] defined in Table. 5.1.

Table 5.1: Complexity of basic operations.

Operation	Computations
Matrix mult. $(m \times n) \cdot (n \times p)$	mnp
Matrix inversion $n \times n$	$\frac{2}{3}n^3$
SVD $m \times n$	$4m^2n + 8mn^2 + 9n^3$

A summary of the complexity of the compared BFs is given in Table. 5.2. The proposed

Table 5.2: Number of computations and memory usage of various closed-form and GSC-form LCMV-BF.

BF	**Computations**	**Memory**
SF-LCMV	$\frac{2}{3}M^3 + \frac{2}{3}P^3 + 2M^2P$ $+MP + P^2$	0
SF-GSC	$9\frac{2}{3}P^3 + 4M^2P + 9MP^2$ $+MP + P^2$	0
SUI-LCMV	$4M^2 + 2MP + 5P^2$ $-6M + 11P + 2$	$2M^2 + 2P^2$
SUD-LCMV	$4M^2 + 2MP + 5P^2$ $-5M + 5P + 1$	$2M^2 + 2P^2$
SUI-GSC	$MP + 5P^2 + 4M$ $+2P - 2$	$2P^2$
SUD-GSC	$2M^2 + 7P^2 - 3MP$ $+5M - 2P - 2$	$2P^2$
CUI-LCMV	$2M^2 + 2MP + 3P^2$ $+2M - 4P + 1$	$2M^2 + 2P^2$
CUD-LCMV	$2M^2 + 2MP + 2P^2$ $+2M - 2P$	$2M^2 + 2P^2$
CUI-GSC	$4M^2 - 5MP + 6P^2$ $+12M - 12P + 6$	$2P^2$
CUD-GSC	$3MP + 3P^2$ $+3M - 4P$	$2P^2$

updating procedures reduce the computational complexity of the SF-LCMV implementation, which is $\mathcal{O}(M^3+M^2P)$, to $\mathcal{O}(M^2+MP)$ while increasing the memory requirement to $\mathcal{O}(M^2+P^2)$. Similarly, regarding the GSC implementation, the updating procedures reduce the computational complexity from $\mathcal{O}(P^3+M^2P+P^2M)$ in the straightforward implementation to $\mathcal{O}(M^2 + MP)$, while increasing the memory requirements to $\mathcal{O}(P^2)$. Please note that the computational complexity of the LCMV and the GSC updating procedures is similar, whereas the memory requirement of the GSC procedures is much lower than its LCMV form counterparts. The number of computations versus the number of sensors while the number

of constraint is fixed to $P = 10$ is depicted in Figs. 5.7 and 5.8 for the closed-form and for the GSC form implementations, respectively. It is evident that the updating procedures impose a lower computational burden. The number of computations versus the number of constraints while the number of sensors is fixed to $M = 20$ is depicted in Figs. 5.9 and 5.10. Again, it is evident that the updating procedures impose a lower computational burden. It is interesting to note that the number of computations of the CUI-GSC and the SUD-GSC is not monotonically increasing with P. This is attributed to the fact that the dimensions of the BM are reversely proportional to the number of constraints. In many applications, the number of constraints can be increased with the number of available sensors. In Figs. 5.11 and 5.12 the computational complexity is depicted versus the number of sensors, while the number of constraints is set to $P = \lfloor \frac{1}{2} M \rfloor$. The complexity reduction is evident from these figures as well.

The overall computational saving is proportional to the BF update rate, whereas the memory complexity is fixed and considerably low. In a dynamically changing network a substantial computational saving is expected. Please notice that even in the case of a single update of the BF, less computations are required when using the proposed updating procedures than in the straightforward recalculation.

5.2.5 Algorithms Summary

In Sec. 5.2.2 we derived the SUI-LCMV, SUI-GSC, CUI-GSC and SUD-GSC algorithms. The derivation was based on matrix algebra, the Woodbury identity, the block matrix inversion formula and the Householder transformation. We use similar methods to derive the rest of the algorithms, namely the incremental or decremental updates of either the number of sensors or the number of constraints for the GSC or the closed-form implementations. We therefore omit the derivation of the rest of the algorithms for brevity. Instead, in the following, we summarize all the proposed low-complexity beamformer updating methods. The sensor updating algorithms SUI-LCMV, SUD-LCMV, SUI-GSC, SUD-GSC and the constraint updating algorithms CUI-LCMV, CUD-LCMV, CUI-GSC, CUD-GSC are summarized in Algs. 2, 3, 6, 7 and Algs. 4, 5, 8, 9, respectively.

Algorithm 2 Summary of the SUI-LCMV procedure

input: $\boldsymbol{C}$, $\boldsymbol{g}$, $\boldsymbol{w}$, $\boldsymbol{\Phi}^{-1}$, $\dot{\sigma}^2$, $\boldsymbol{Q}^{-1}$, $\dot{\boldsymbol{\phi}}$, $\dot{\boldsymbol{c}}$

output: $\dot{\boldsymbol{C}}$, $\dot{\boldsymbol{w}}$, $\dot{\boldsymbol{\Phi}}^{-1}$, $\dot{\boldsymbol{Q}}^{-1}$

begin

$$\dot{\boldsymbol{C}} = \left[\begin{array}{c|c} \boldsymbol{C}^H & \dot{\boldsymbol{c}} \end{array}\right]^H$$

$$\dot{\epsilon} = \left(\dot{\sigma}^2 - \dot{\boldsymbol{\phi}}^H \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}}\right)^{-1}$$

$$\dot{\boldsymbol{q}} = \boldsymbol{C}^H \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}} - \dot{\boldsymbol{c}}$$

$$\dot{\boldsymbol{\Phi}}^{-1} = \left[\begin{array}{c|c} \boldsymbol{\Phi}^{-1} + \dot{\epsilon} \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}} \dot{\boldsymbol{\phi}}^H \boldsymbol{\Phi}^{-1} & -\dot{\epsilon} \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}} \\ \hline -\dot{\epsilon} \dot{\boldsymbol{\phi}}^H \boldsymbol{\Phi}^{-1} & \dot{\epsilon} \end{array}\right]$$

$$\dot{\boldsymbol{Q}}^{-1} = \boldsymbol{Q}^{-1} - \frac{\boldsymbol{Q}^{-1} \dot{\boldsymbol{q}} \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}}}$$

$$\dot{w}_M = -\dot{\epsilon} \left(\dot{\boldsymbol{\phi}}^H \boldsymbol{\Phi}^{-1} \boldsymbol{C} \boldsymbol{Q}^{-1} - \dot{\boldsymbol{c}}^H \boldsymbol{Q}^{-1}\right) \left(\boldsymbol{I}_{P \times P} - \frac{\dot{\boldsymbol{q}} \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}}}\right) \boldsymbol{g}$$

$$\dot{\boldsymbol{\Delta}} \dot{\boldsymbol{w}} = -\frac{\dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \boldsymbol{g}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}}} \boldsymbol{\Phi}^{-1} \boldsymbol{C} \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}} - \dot{w}_M \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}}$$

$$\dot{\boldsymbol{w}} = \left[\begin{array}{c} \boldsymbol{w} + \dot{\boldsymbol{\Delta}} \dot{\boldsymbol{w}} \\ \hline \dot{w}_M \end{array}\right]$$

end

Algorithm 3 Summary of the SUD-LCMV procedure

input: $\dot{\boldsymbol{C}}$, $\boldsymbol{g}$, $\dot{\boldsymbol{w}}$, $\dot{\boldsymbol{\Phi}}^{-1}$, $\dot{\boldsymbol{Q}}^{-1}$, $\dot{\sigma}^2$, $\dot{\boldsymbol{\phi}}$

output: $\boldsymbol{C}$, $\boldsymbol{w}$, $\boldsymbol{\Phi}$, $\boldsymbol{Q}^{-1}$

begin

$$\boldsymbol{C} = \bar{\boldsymbol{I}}_M^H \dot{\boldsymbol{C}}$$

$$\boldsymbol{\Phi}^{-1} = \bar{\boldsymbol{I}}_M^H \dot{\boldsymbol{\Phi}}^{-1} \bar{\boldsymbol{I}}_M - \frac{\bar{\boldsymbol{I}}_M^H \dot{\boldsymbol{\Phi}}^{-1} \dot{\boldsymbol{\imath}}_M \dot{\boldsymbol{\imath}}_M^H \dot{\boldsymbol{\Phi}}^{-1} \bar{\boldsymbol{I}}_M}{\dot{\boldsymbol{\imath}}_M^H \dot{\boldsymbol{\Phi}}^{-1} \dot{\boldsymbol{\imath}}_M}$$

$$\dot{\epsilon} = \left(\dot{\sigma}^2 - \dot{\boldsymbol{\phi}}^H \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}}\right)^{-1}$$

$$\dot{\boldsymbol{q}} = \boldsymbol{C}^H \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}} - \dot{\boldsymbol{c}}$$

$$\boldsymbol{Q}^{-1} = \dot{\boldsymbol{Q}}^{-1} + \frac{\dot{\boldsymbol{Q}}^{-1} \dot{\boldsymbol{q}} \dot{\boldsymbol{q}}^H \dot{\boldsymbol{Q}}^{-1}}{\dot{\epsilon}^{-1} - \dot{\boldsymbol{q}}^H \dot{\boldsymbol{Q}}^{-1} \dot{\boldsymbol{q}}}$$

$$\dot{w}_M = -\dot{\epsilon} \left(\dot{\boldsymbol{\phi}}^H \boldsymbol{\Phi}^{-1} \boldsymbol{C} \boldsymbol{Q}^{-1} - \dot{\boldsymbol{c}}^H \boldsymbol{Q}^{-1}\right) \left(\boldsymbol{I}_{P \times P} - \frac{\dot{\boldsymbol{q}} \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}}}\right) \boldsymbol{g}$$

$$\dot{\boldsymbol{\Delta}} \dot{\boldsymbol{w}} = -\frac{\dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \boldsymbol{g}}{\dot{\epsilon}^{-1} + \dot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}}} \boldsymbol{\Phi}^{-1} \boldsymbol{C} \boldsymbol{Q}^{-1} \dot{\boldsymbol{q}} - \dot{w}_M \boldsymbol{\Phi}^{-1} \dot{\boldsymbol{\phi}}$$

$$\boldsymbol{w} = \bar{\boldsymbol{I}}_M^H \dot{\boldsymbol{w}} - \dot{\boldsymbol{\Delta}} \dot{\boldsymbol{w}}$$

end

Algorithm 4 Summary of the CUI-LCMV procedure

input: $\boldsymbol{C}$, $\boldsymbol{g}$, $\boldsymbol{w}$, $\boldsymbol{\Phi}^{-1}$, $\boldsymbol{Q}^{-1}$, $\ddot{\boldsymbol{c}}$, $\ddot{g}$

output: $\ddot{\boldsymbol{C}}$, $\ddot{\boldsymbol{g}}$, $\ddot{\boldsymbol{w}}$, $\ddot{\boldsymbol{Q}}^{-1}$

begin

$$\ddot{\boldsymbol{C}} = \left[\begin{array}{c|c} \boldsymbol{C} & \ddot{\boldsymbol{c}} \end{array}\right]$$

$$\ddot{\boldsymbol{g}} = \left[\begin{array}{c|c} \boldsymbol{g}^H & \ddot{g}^* \end{array}\right]^H$$

$$\ddot{\boldsymbol{q}} = \boldsymbol{C}^H \boldsymbol{\Phi}^{-1} \ddot{\boldsymbol{c}}$$

$$\ddot{\eta} = \left(\ddot{\boldsymbol{c}}^H \boldsymbol{\Phi}^{-1} \ddot{\boldsymbol{c}} - \ddot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \ddot{\boldsymbol{q}}\right)^{-1}$$

$$\ddot{\boldsymbol{Q}}^{-1} = \left[\begin{array}{c|c} \boldsymbol{Q}^{-1} + \ddot{\eta} \boldsymbol{Q}^{-1} \ddot{\boldsymbol{q}} \ddot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} & -\ddot{\eta} \boldsymbol{Q}^{-1} \ddot{\boldsymbol{q}} \\ \hline -\ddot{\eta} \ddot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} & \ddot{\eta} \end{array}\right]$$

$$\ddot{\boldsymbol{\Delta w}} = \ddot{\eta} \left(\ddot{g} - \ddot{\boldsymbol{c}}^H \boldsymbol{w}\right) \boldsymbol{\Phi}^{-1} \left(\boldsymbol{I} - \boldsymbol{C} \boldsymbol{Q}^{-1} \boldsymbol{C}^H \boldsymbol{\Phi}^{-1}\right) \ddot{\boldsymbol{c}}$$

$$\ddot{\boldsymbol{w}} = \boldsymbol{w} + \ddot{\boldsymbol{\Delta w}}$$

end

Algorithm 5 Summary of the CUD-LCMV procedure

input: $\ddot{\boldsymbol{C}}$, $\ddot{\boldsymbol{g}}$, $\ddot{\boldsymbol{w}}$, $\boldsymbol{\Phi}^{-1}$, $\ddot{\boldsymbol{Q}}^{-1}$

output: $\boldsymbol{C}$, $\boldsymbol{g}$, $\boldsymbol{w}$, $\boldsymbol{Q}^{-1}$

begin

$$\boldsymbol{C} = \ddot{\boldsymbol{C}} \bar{\boldsymbol{I}}_P$$

$$\boldsymbol{g} = \bar{\boldsymbol{I}}_P^H \ddot{\boldsymbol{g}}$$

$$\boldsymbol{Q}^{-1} = \bar{\boldsymbol{I}}_P^H \ddot{\boldsymbol{Q}}^{-1} \bar{\boldsymbol{I}}_P - \frac{\bar{\boldsymbol{I}}_P^H \ddot{\boldsymbol{Q}}^{-1} \boldsymbol{i}_P \boldsymbol{i}_P^H \ddot{\boldsymbol{Q}}^{-1} \bar{\boldsymbol{I}}_P}{\boldsymbol{i}_P^H \ddot{\boldsymbol{Q}}^{-1} \boldsymbol{i}_P}$$

$$\ddot{\boldsymbol{q}} = \boldsymbol{C}^H \boldsymbol{\Phi}^{-1} \ddot{\boldsymbol{c}}$$

$$\ddot{\eta} = \left(\ddot{\boldsymbol{c}}^H \boldsymbol{\Phi}^{-1} \ddot{\boldsymbol{c}} - \ddot{\boldsymbol{q}}^H \boldsymbol{Q}^{-1} \ddot{\boldsymbol{q}}\right)^{-1}$$

$$\ddot{\boldsymbol{\Delta w}} = \ddot{\eta} \left(\ddot{g} - \ddot{\boldsymbol{c}}^H \boldsymbol{w}\right) \boldsymbol{\Phi}^{-1} \left(\boldsymbol{I} - \boldsymbol{C} \boldsymbol{Q}^{-1} \boldsymbol{C}^H \boldsymbol{\Phi}^{-1}\right) \ddot{\boldsymbol{c}}$$

$$\boldsymbol{w} = \ddot{\boldsymbol{w}} - \ddot{\boldsymbol{\Delta w}}$$

end

Algorithm 6 Summary of the SUI-GSC procedure

input: $\boldsymbol{C}$, $\boldsymbol{g}$, $\boldsymbol{w}_{\parallel}$, $\boldsymbol{R}^{-1}$, $\boldsymbol{B}$, $\dot{\boldsymbol{c}}$

output: $\dot{\boldsymbol{C}}$, $\dot{\boldsymbol{w}}_{\parallel}$, $\dot{\boldsymbol{R}}^{-1}$, $\dot{\boldsymbol{B}}$

begin

$$\dot{\boldsymbol{C}} = \left[\begin{array}{c|c} \boldsymbol{C}^H & \dot{\boldsymbol{c}} \end{array} \right]^H$$

$$\dot{\boldsymbol{R}}^{-1} = \boldsymbol{R}^{-1} - \frac{\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}\dot{\boldsymbol{c}}^H\dot{\boldsymbol{R}}^{-1}}{1+\dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}}$$

$$\dot{w}_{\parallel M} = \frac{\dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\boldsymbol{g}}{1+\dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}}$$

$$\dot{\boldsymbol{\Delta w}}_{\parallel} = -\dot{w}_{\parallel M}\boldsymbol{C}\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}$$

$$\dot{\boldsymbol{w}}_{\parallel} = \left[\begin{array}{c|c} \left(\boldsymbol{w}_{\parallel} + \dot{\boldsymbol{\Delta w}}_{\parallel}\right)^T & \dot{w}_{\parallel M} \end{array} \right]^T$$

$$\dot{\boldsymbol{\Delta b}} = \left[\begin{array}{c|c} \dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\boldsymbol{C}^H & -1 \end{array} \right]^H / \| \left[\begin{array}{c|c} \dot{\boldsymbol{c}}^H\boldsymbol{R}^{-1}\boldsymbol{C}^H & -1 \end{array} \right] \|$$

$$\dot{\boldsymbol{B}} = \left[\begin{array}{c|c} \boldsymbol{B} & \\ \boldsymbol{0}_{1\times M-1-P} & \dot{\boldsymbol{\Delta b}} \end{array} \right]$$

end

Algorithm 7 Summary of the SUD-GSC procedure

input: $\dot{\boldsymbol{C}}$, $\boldsymbol{g}$, $\dot{\boldsymbol{w}}_{\parallel}$, $\dot{\boldsymbol{R}}^{-1}$, $\dot{\boldsymbol{B}}$

output: $\boldsymbol{C}$, $\boldsymbol{w}_{\parallel}$, $\boldsymbol{R}^{-1}$, $\boldsymbol{B}$

begin

$$\boldsymbol{C} = \bar{\boldsymbol{I}}_M^H\dot{\boldsymbol{C}}$$

$$\boldsymbol{R}^{-1} = \dot{\boldsymbol{R}}^{-1} + \frac{\dot{\boldsymbol{R}}^{-1}\dot{\boldsymbol{c}}\dot{\boldsymbol{c}}^H\dot{\boldsymbol{R}}^{-1}}{1-\dot{\boldsymbol{c}}^H\dot{\boldsymbol{R}}^{-1}\dot{\boldsymbol{c}}}$$

$$\dot{w}_{\parallel M} = \boldsymbol{i}_M^H\dot{\boldsymbol{w}}_{\parallel}$$

$$\dot{\boldsymbol{\Delta w}}_{\parallel} = -\dot{w}_{\parallel M}\boldsymbol{C}\boldsymbol{R}^{-1}\dot{\boldsymbol{c}}$$

$$\boldsymbol{w}_{\parallel} = \bar{\boldsymbol{I}}_M^H\dot{\boldsymbol{w}}_{\parallel} - \dot{\boldsymbol{\Delta w}}_{\parallel}$$

$$\dot{\boldsymbol{\chi}} = \dot{\boldsymbol{B}}^H\boldsymbol{i}_M + \|\dot{\boldsymbol{B}}^H\boldsymbol{i}_M\| \exp\left(-j\angle\dot{B}_{M,M-P}\right)\boldsymbol{i}_{M-P}$$

$$\boldsymbol{B} = \bar{\boldsymbol{I}}_M^H\dot{\boldsymbol{B}}\left(\boldsymbol{I} - \frac{2\dot{\boldsymbol{\chi}}\dot{\boldsymbol{\chi}}^H}{\|\dot{\boldsymbol{\chi}}\|^2}\right)\bar{\boldsymbol{I}}_{M-P}$$

end

Algorithm 8 Summary of the CUI-GSC procedure

input: $\boldsymbol{C}$, $\boldsymbol{g}$, $\boldsymbol{w}_{\parallel}$, $\boldsymbol{R}^{-1}$, $\boldsymbol{B}$, $\ddot{\boldsymbol{c}}$, $\ddot{g}$

output: $\ddot{\boldsymbol{C}}$, $\ddot{\boldsymbol{g}}$, $\ddot{\boldsymbol{w}}_{\parallel}$, $\ddot{\boldsymbol{R}}^{-1}$, $\ddot{\boldsymbol{B}}$

begin

$$\ddot{\boldsymbol{C}} = \left[\begin{array}{c|c} \boldsymbol{C} & \ddot{\boldsymbol{c}} \end{array}\right]$$

$$\ddot{\boldsymbol{g}} = \left[\begin{array}{c|c} \boldsymbol{g}^H & \ddot{g}^* \end{array}\right]^H$$

$$\ddot{\boldsymbol{r}} = \boldsymbol{C}^H \ddot{\boldsymbol{c}}$$

$$\ddot{\rho} = \left(\|\ddot{\boldsymbol{c}}\|^2 - \ddot{\boldsymbol{r}}^H \boldsymbol{R}^{-1} \ddot{\boldsymbol{r}}\right)^{-1}$$

$$\ddot{\boldsymbol{R}}^{-1} = \left[\begin{array}{c|c} \boldsymbol{R}^{-1} + \ddot{\rho}\boldsymbol{R}^{-1}\ddot{\boldsymbol{r}}\ddot{\boldsymbol{r}}^H\boldsymbol{R}^{-1} & -\ddot{\rho}\boldsymbol{R}^{-1}\ddot{\boldsymbol{r}} \\ \hline -\ddot{\rho}\ddot{\boldsymbol{r}}^H\boldsymbol{R}^{-1} & \ddot{\rho} \end{array}\right]$$

$$\ddot{\boldsymbol{\Delta w}}_{\parallel} = \ddot{\rho}\left(\ddot{g} - \ddot{\boldsymbol{c}}^H \boldsymbol{w}_{\parallel}\right)\left(\boldsymbol{I} - \boldsymbol{C}\boldsymbol{R}^{-1}\boldsymbol{C}^H\right)\ddot{\boldsymbol{c}}$$

$$\ddot{\boldsymbol{w}}_{\parallel} = \boldsymbol{w}_{\parallel} + \ddot{\boldsymbol{\Delta w}}_{\parallel}$$

$$\ddot{\boldsymbol{b}} = \boldsymbol{B}^H \ddot{\boldsymbol{c}}$$

$$\ddot{\boldsymbol{\chi}} = \frac{\ddot{\boldsymbol{b}}}{\|\ddot{\boldsymbol{b}}\|} + \exp\left(j\angle\ddot{b}_{M-P+1}\right)\boldsymbol{i}_{M-P+1}$$

$$\ddot{\boldsymbol{B}} = \boldsymbol{B}\left(\boldsymbol{I}_{(M-P+1)\times(M-P+1)} - \frac{2\ddot{\boldsymbol{\chi}}\ddot{\boldsymbol{\chi}}^H}{\|\ddot{\boldsymbol{\chi}}\|^2}\right)\bar{\boldsymbol{I}}_{M-P+1}$$

end

Algorithm 9 Summary of the CUD-GSC procedure

input: $\ddot{\boldsymbol{C}}$, $\ddot{\boldsymbol{g}}$, $\ddot{\boldsymbol{w}}_{\parallel}$, $\ddot{\boldsymbol{R}}^{-1}$, $\ddot{\boldsymbol{B}}$

output: $\boldsymbol{C}$, $\boldsymbol{g}$, $\boldsymbol{w}_{\parallel}$, $\boldsymbol{R}^{-1}$, $\boldsymbol{B}$

begin

$$\boldsymbol{C} = \ddot{\boldsymbol{C}}\bar{\boldsymbol{I}}_P$$

$$\boldsymbol{g} = \bar{\boldsymbol{I}}_P^H \ddot{\boldsymbol{g}}$$

$$\boldsymbol{R}^{-1} = \bar{\boldsymbol{I}}_P^H \ddot{\boldsymbol{R}}^{-1}\bar{\boldsymbol{I}}_P - \frac{\bar{\boldsymbol{I}}_P^H\ddot{\boldsymbol{R}}^{-1}\boldsymbol{i}_P\boldsymbol{i}_P^H\ddot{\boldsymbol{R}}^{-1}\bar{\boldsymbol{I}}_P}{\boldsymbol{i}_P^H\ddot{\boldsymbol{R}}^{-1}\boldsymbol{i}_P}$$

$$\boldsymbol{\Delta w}_{\parallel} = \|\ddot{\boldsymbol{C}}\ddot{\boldsymbol{R}}^{-1}\boldsymbol{i}_P\|^{-2}\ddot{\boldsymbol{C}}\ddot{\boldsymbol{R}}^{-1}\boldsymbol{i}_P\left(\ddot{\boldsymbol{C}}\ddot{\boldsymbol{R}}^{-1}\boldsymbol{i}_P\right)^H \ddot{\boldsymbol{w}}_{\parallel}$$

$$\boldsymbol{w}_{\parallel} = \ddot{\boldsymbol{w}}_{\parallel} - \boldsymbol{\Delta w}_{\parallel}$$

$$\boldsymbol{B} = \left[\begin{array}{c|c} \ddot{\boldsymbol{B}} & \frac{\ddot{\boldsymbol{C}}\ddot{\boldsymbol{R}}^{-1}\boldsymbol{i}_P}{\|\ddot{\boldsymbol{C}}\ddot{\boldsymbol{R}}^{-1}\boldsymbol{i}_P\|} \end{array}\right]$$

end

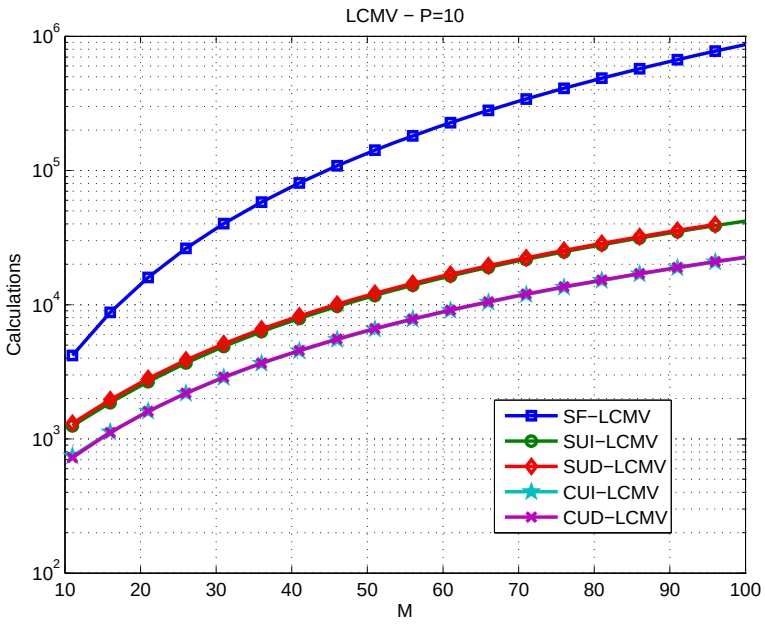


Figure 5.7: Number of computations vs. M for LCMV-BFs with $P = 10$.

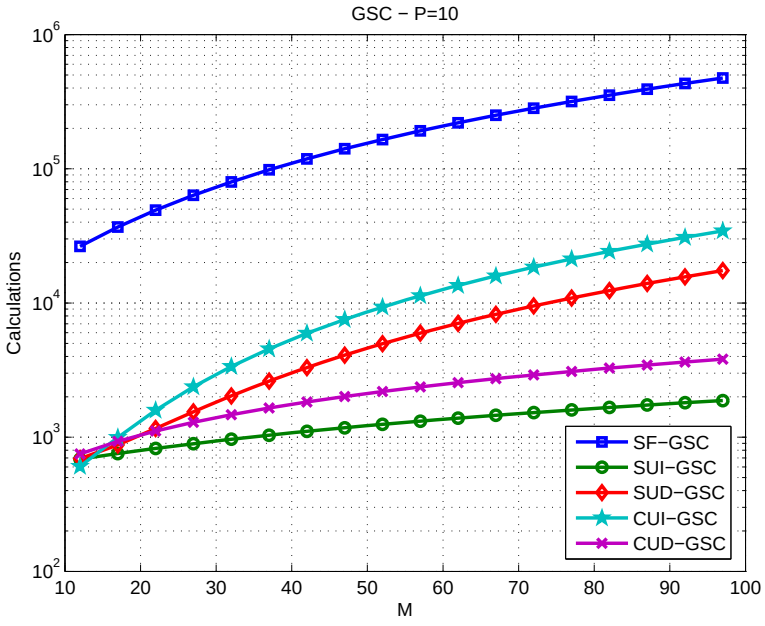


Figure 5.8: Number of computations vs. M for GSC-BFs with $P = 10$.

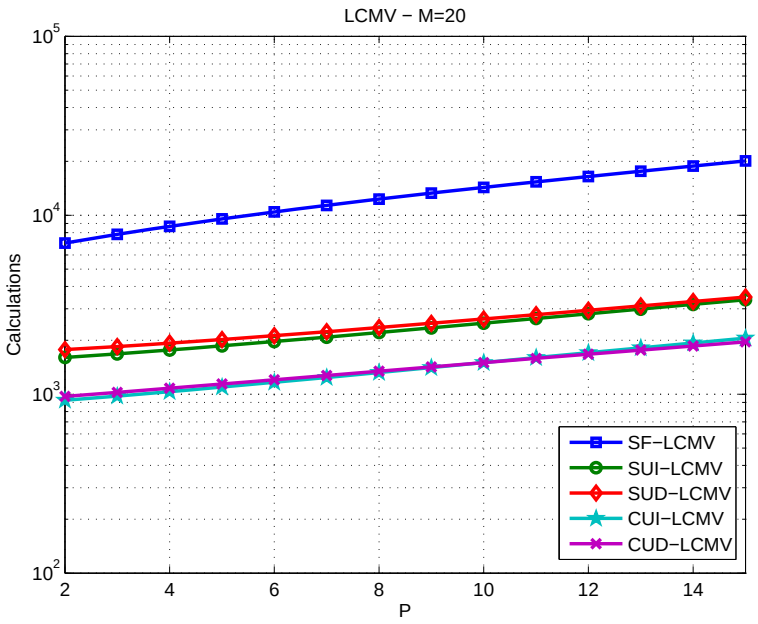


Figure 5.9: Number of computations vs. P for LCMV-BFs with $M = 20$.

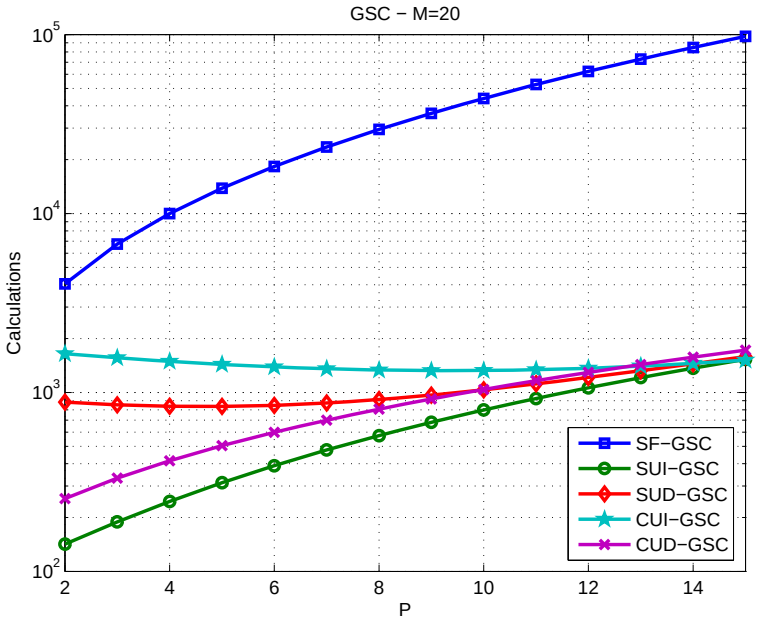


Figure 5.10: Number of computations vs. P for GSC-BFs with $M = 20$.

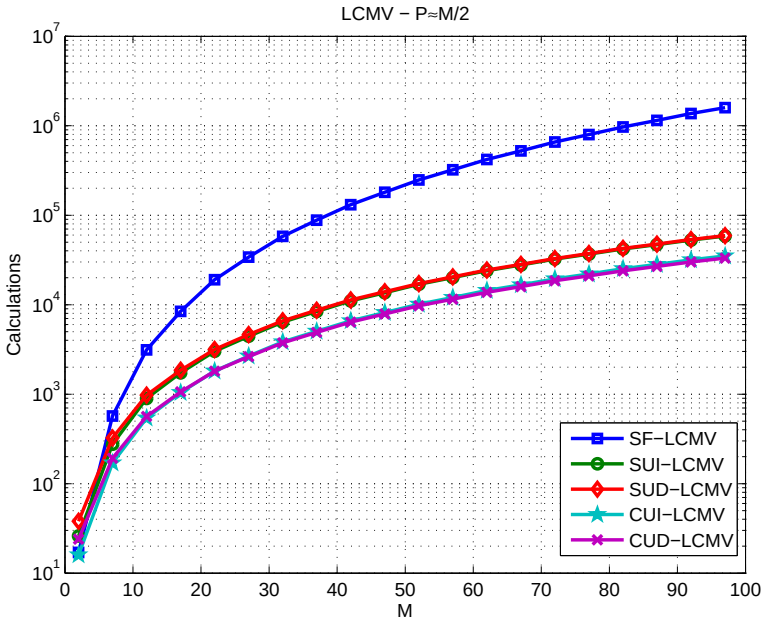


Figure 5.11: Number of computations vs. M for LCMV-BFs with $P = \lfloor \frac{1}{2} M \rfloor$.

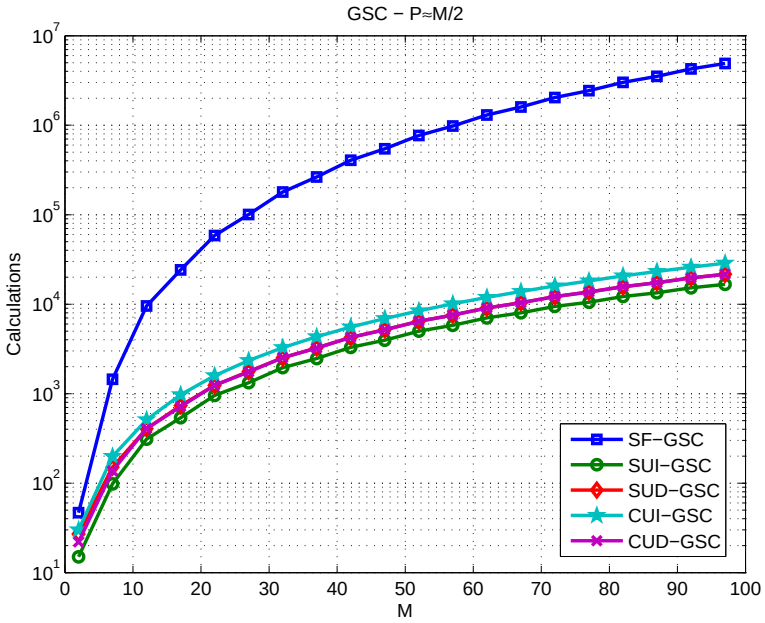


Figure 5.12: Number of computations vs. M for GSC-BFs with $P = \lfloor \frac{1}{2} M \rfloor$.

5.2.6 Conclusions

Procedures for adding/removing active sensors or constraints to/from an existing LCMV-BF have been derived. Different procedures were derived for both closed-form and GSC-form implementations. These procedures use the information of the former BF and save calculations, at the expense of some memory requirements. The computational burden of the proposed procedures was analyzed and compared with the computational burden of their corresponding straightforward BFs recalculation. It is evident from the comparison that the number of computations in the proposed procedures is much lower than in straightforward calculation, while the increase in the memory complexity is considerably low. The proposed procedures are beneficial in sensor network applications, where the dynamics of the network and of the environment require frequent updates of the BF, whereas the computational capability is often limited.

Chapter 6

Summary and future research directions

6.1 Summary

Technology advances in hardware and communication have made the vision of WASN feasible and within grasp. Sensor networks spread over vast environments hold potential for superior performance to classical condensed microphone arrays. WASNs raise several new challenges. This dissertation addressed some of the challenges and propose possible solutions. The contributions of this dissertation are listed below:

1. Distributed versions of classical beamforming algorithms were developed. We consider the general scenario of P desired and interfering speakers received by M microphones, clustered in an N nodes. All sources propagate in a noisy and reverberant environment. A distributed GSC algorithm was developed for this challenging scenario. The proposed BF is time-recursive and its convergence to the optimal centralized LCMV-BF is guaranteed. It requires only $N+P$ communication channels. Considering the special case of a single speaker, we derived a distributed BF which further reduces the required communication-bandwidth to N communication-channels. For a binaural hearing aid system we derived an iterative MVDR-BF for enhancing a single desired speaker while maintaining the spatial cues. We also addressed the problem of sampling rate offsets between nodes. An algorithm for estimating the offsets and compensating for them was derived based on the noise stationarity.

2. A novel framework for evaluating beamformers in WASN is derived. In classical array

processing the layout of the array is usually pre-determined to fit the problem at hand, contrary to WASN where the layout can be random and dynamic. The performance of the SDW-MWF BF with randomly located microphones in a reverberant environment, contaminated by either coherent or diffuse noise fields, was analyzed. The analysis may serve as a design guideline for determining the number of nodes/microphones required in order to meet a desired performance level.

3. WASNs comprise multiple nodes and microphones. Price and battery-life constraints limit the allowed computational complexity of each node. Frequent updates of the BF are required due to the dynamics of the scenario. Therefore, efficient implementation of the local BFs becomes necessary. Hence, we addressed the problem of reducing the complexity of applying an LCMV-BF, without sacrificing performance. The BM filters, by far, dominate the computation burden in applying the GSC-BF. We propose a sparse BM structure which reduces computational complexity significantly from $M \times (M-P)$ to $P \times (M-P)$. Considering environment dynamics, efficient methods for modifying the BF were developed, including addition/removal of sensors and/or constraints.

4. Additionaly, we developed a novel criterion for speakers extraction in a noisy reverberant environment. The proposed method is a generalization of the SDW-MWF to multiple sources. We proved that two special cases of the new method are the MWF and the LCMV BFs. Furthermore, considering a dynamic scenario in which the speakers move around the room, we derived an extension to the eigen-spaces LCMV-BF with source tracking capability, which is based on the PASTd algorithm.

6.2 Future resarch directions

The following topics and directions are left to future research:

Oracle: The derived algorithms assume the existence of an "oracle", able to mark the activity patterns of the various sources. We will examine various approaches for estimating the sources activity, and hopefully find a suitable and robust method for obtaining such an oracle.

Robustness and dynamics: Based on the centralized subspace tracking BF mentioned above, we will introduce tracking capabilities to the DGSC.

Sub-optimal, short latency solutions: The derived distributed BF, which converges to the optimal solution, involves applying long filters in correspondence to the length of the RIR. We will explore sub-optimal solutions which require a shorter latency (a combination of linear filtering and non-linear post-filtering), and therefore are more adequate for real-time applications.

Combined noise reduction-echo cancelation: Thus far we focused on the noise reduction problem. However, other issues should be addressed as well. Ad hoc teleconferencing systems require cancelling of the far-end side emitted by speakers in the conference room. Distributed algorithms which combine noise reduction and echo cancellation could be developed.

Dereverberation: Poor acoustic design of the conference rooms, could result in deteriorated intelligibility. This problem can be alleviated by introducing dereveberation techniques to the proposed distributed algorithm. The spatial distribution of the microphones might be beneficial in estimating the required room acoustic parameters.

MSDW-MWF: The proposed beamforming criterion introduces more degrees of freedom in designing a desired response. We intend to examine their ability for controlling the BF sensitivity and derive a robust distributed BF for WASN.

Bibliography

[1] D. Estrin, G. Pottie, and M. Srivastava, "Instrumenting the world with wireless sensor networks," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2033–2036, May 2001.

[2] J. B. Predd, S. R. Kulkarni, and H. V. Poor, "Distributed learning in wireless sensor networks," *IEEE Signal Processing Magazine*, vol. 23, no. 4, pp. 56–69, Jul. 2006.

[3] D. Culler, D. Estrin, and M. Srivastava, "Overview of sensor networks," *Computer*, vol. 37, no. 8, pp. 41–49, Aug. 2004.

[4] H. Ochiai, P. Mitran, H.V. Poor, and V. Tarokh, "Collaborative beamforming for distributed wireless ad hoc sensor networks," *IEEE Transactions on Signal Processing*, vol. 53, no. 11, pp. 4110–4124, Nov. 2005.

[5] M.F.A Ahmed and S.A Vorobyov, "Collaborative beamforming for wireless sensor networks with Gaussian distributed sensor nodes," *IEEE Transactions on Wireless Communications*, vol. 8, no. 2, pp. 638–643, Feb. 2009.

[6] Y. T. Lo, "A mathematical theory of antenna arrays with randomly spaced elements," *IEEE Trans. on Antennas and Propagation*, vol. 12, pp. 257–268, May 1964.

[7] J.F. Cardoso, "Blind signal separation: Statistical principles," *Proc. of the IEEE*, vol. 86, no. 10, pp. 2009–2025, Oct. 1998.

[8] P. Comon, "Independent component analysis: A new concept?," *Signal Processing*, vol. 36, no. 3, pp. 287–314, Apr. 1994.

[9] L. Parra and C. Spence, "Convolutive blind separation of non-stationary sources," *IEEE Trans. Speech and Audio Processing*, vol. 8, no. 3, pp. 320–327, May 2000.

[10] A. Hyviirinen, J. Karhunen, and E. Oja, *ndependent Component Analysis*, Wiley & Sons, Inc., New York, 2001.

[11] L. Molgedey and H. G. Schuster, "Separation of a mixture of independent signals using time delayed correlations," *Phys. Rev. Lett.*, vol. 72, no. 23, pp. 3634–3637, June 1994.

[12] J. F. Cardoso, "Eigen-structure of the 4th-order cumulant tensor with application to the blind source separation problem," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 2109–2112, May 1989.

[13] S. Amari, A. Chichocki, and H. H. Yang, "Blind signal separation and extraction: Neural and information-theoretic approaches," *Unsupervised Adaptive Filtering*, vol. 1, 2000.

[14] H. Wu and J.C. Principe, "A unifying criterion for blind source separation and decorrelation: Simultaneous diagonalization of correlation matrices," *Proc. IEEE Workshop on Neural Networks for Signal Processing (NNSP)*, pp. 496–508, Sep. 1997.

[15] M. Z. Ikram and D. R. Morgan, "Exploring permutation inconsistency in blind separation of speech signals in a reverberant environment," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 2, pp. 1041–1044, June 2000.

[16] M. Souden and Z. Liu, "Optimal joint linear acoustic echo cancelation and blind source separation in the presence of loudspeaker nonlinearity," in *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2009, pp. 117–120.

[17] H. Kameoka, M. Sato, T. Ono, N. Ono, and S. Sagayama, "Blind separation of infinitely many sparse sources," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2012.

[18] N. Ono, K. Miyamoto, H. Kameoka, J. Le Roux, Y. Uchiyama, E. Tsunoo, T. Nishimoto, and S. Sagayama, "Harmonic and percussive sound separation and its application to MIR-related tasks," in *Advances in Music Information Retrieval*, pp. 213–236. Springer, 2010.

[19] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures exploiting nongaussianity, nonwhiteness, and nonstationarity," in *Proc. Int.*

Workshop on Acoustic Echo and Noise Control (IWAENC), Kyoto, Japan, Sep. 2003, pp. 223–226.

[20] H. Buchner, R. Aichner, and W. Kellermann, "Blind source separation for convolutive mixtures: A unified treatment," in *Audio Signal Processing for Next-Generation Multimedia Communication Systems*, Y. Huang and J. Benesty, Eds., pp. 255–293. Kluwer Academic Publ., Boston, 2004.

[21] H. Buchner, R. Aichner, and W. Kellermann, "Trinicon: A versatile framework for multichannel blind signal processing," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Montreal, Canada, May 2004, pp. 889–892.

[22] H. Buchner, R. Aichner, and W. Kellermann, "A generalization of blind source separation algorithms for convolutive mixtures based on second-order statistics," *IEEE Transactions on Signal Processing*, vol. 13, no. 1, pp. 120–134, Jan. 2005.

[23] H. Buchner, R. Aichner, and W. Kellermann, "Trinicon-based blind system identification with application to multiple-source localization and separation," in *Blind Speech Separation*, S. Makino, T. w. Lee, and S. Sawada, Eds., pp. 101–147. Springer-Verlag, BerlinlHeidelberg, 2007.

[24] E. Jan and J. Flanagan, "Microphone arrays for speech processing," *Int. Symposium on Signals, Systems, and Electronics (ISSSE)*, pp. 373–376, Oct. 1995.

[25] B. D. Van Veen and K. M. Buckley, "Beamforming: A versatile approach to spatial filtering," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 5, no. 2, pp. 4–24, Apr. 1988.

[26] S. Gannot and I. Cohen, "Adaptive beamforming and postfiltering," in *Springer Handbook of Speech Processing*, J. Benesty, M. Mohan Sondhi, and Yiteng (Arden) Huang, Eds., chapter 47, pp. 945–978. Springer, 2007.

[27] Henry Cox, Robert M. Zeskind, and Mark M. Owen, "Robust adaptive beamforming," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 35, no. 10, Oct. 1987.

[28] S. Doclo and M. Moonen, "GSVD-based optimal filtering for multi-microphone speech enhancement," in *Microphone arrays: Signal processing techniques and applications*, pp. 111–132. Springer, 2001.

[29] S. Doclo and M. Moonen, "GSVD-Based optimal filtering for single and multimicrophone speech enhancement," *IEEE Transactions on Signal Processing*, vol. 50, no. 9, pp. 2230–2244, Sep. 2002.

[30] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Processing*, vol. 84, no. 12, pp. 2367–2387, Dec. 2004.

[31] S. Doclo, A. Spriet, J. Wouters, and M. Moonen, "Speech distortion weighted multichannel Wiener filtering techniques for noise reduction," in *Speech Enhancement*, J. Benesty, S. Makino, and J. Chen, Eds., pp. 199–228. Springer, 2005.

[32] S. Doclo, S. Gannot, M. Moonen, and A. Spriet, "Acoustic beamforming for hearing aid applications," in *Handbook on Array Processing and Sensor Networks*, S. Haykin and K. Ray Liu, Eds. Wiley, 2008.

[33] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proc. IEEE*, vol. 57, no. 8, pp. 1408–1418, Aug. 1969.

[34] O. L. Frost III, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE*, vol. 60, no. 8, pp. 926–935, Aug. 1972.

[35] L. J. Griffiths and C. W. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. on Antennas and Propagation*, vol. 30, pp. 27–34, Jan. 1982.

[36] M. Er and A. Cantoni, "Derivative constraints for broad-band element space antenna array processors," *IEEE Trans. Acoust., Speech, Signal Processing*, vol. 31, no. 6, pp. 1378–1393, Dec. 1983.

[37] B. R. Breed and J. Strauss, "A short proof of the equivalence of LCMV and GSC beamforming," *IEEE Signal Processing Lett.*, vol. 9, no. 6, pp. 168–169, Jun. 2002.

[38] S. Affes and Y. Grenier, "A signal subspace tracking algorithm for microphone array processing of speech," *IEEE Trans. Speech and Audio Processing*, vol. 5, no. 5, pp. 425–437, Sep. 1997.

[39] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Transactions on Signal Processing*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[40] Y. Ephraim and H.L. Van Trees, "A signal subspace approach for speech enhancement," *IEEE Trans. Speech and Audio Processing*, vol. 3, no. 4, pp. 251–266, July 1995.

[41] Y. Hu and P.C. Loizou, "A generalized subspace approach for enhancing speech corrupted by colored noise," *IEEE Trans. Speech and Audio Processing*, vol. 11, no. 4, pp. 334–341, July 2003.

[42] S. Gazor, S. Affes, and Y. Grenier, "Robust adaptive beamforming via target tracking," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1589–1593, June 1996.

[43] B. Yang, "Projection approximation subspace tracking," *IEEE Transactions on Signal Processing*, vol. 43, no. 1, pp. 95–107, Jan. 1995.

[44] S. Gazor, S. Affes, and Y. Grenier, "Wideband multi-source beamforming with adaptive array location calibration and direction finding," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 1904–1907, May 1995.

[45] E. Warsitz, A. Krueger, and R. Haeb-Umbach, "Speech enhancement with a new generalized eigenvector blocking matrix for application in generalized sidelobe canceller," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 73–76, Apr. 2008.

[46] S. Affes, S. Gazor, and Y. Grenier, "An algorithm for multi-source beamforming and multi-target tracking," *IEEE Transactions on Signal Processing*, vol. 44, no. 6, pp. 1512–1522, Jun. 1996.

[47] F. Asano, S. Hayamizu, T. Yamada, and S. Nakamura, "Speech enhancement based on the subspace method," vol. 8, no. 5, pp. 497–507, Sep. 2000.

[48] R. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. on Antennas and Propagation*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[49] J. Benesty, J. Chen, Y. Huang, and J. Dmochowski, "On microphone-array beamforming from a MIMO acoustic signal processing perspective," *IEEE Trans. Audio, Speech and Language Processing*, vol. 15, no. 3, pp. 1053–1065, Mar. 2007.

[50] G. Reuven, S. Gannot, and I. Cohen, "Dual-source transfer-function generalized sidelobe canceller," *IEEE Trans. Audio, Speech and Language Processing*, vol. 16, no. 4, pp. 711–727, May 2008.

[51] S. Markovich-Golan, S. Gannot, and I. Cohen, "Multichannel eigenspace beamforming in a reverberant noisy environment with multiple interfering speech signals," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 6, pp. 1071–1086, Aug. 2009.

[52] S. S. Pradhan, J. Kusuma, and K. Ramchandran, "Distributed compression in a dense microsensor network," *IEEE Signal Processing Magazine*, pp. 51–60, Mar. 2002.

[53] C. Chong and S. P. Kumar, "Sensor networks: Evolution, opportunities and challenges," *Proc. of the IEEE*, vol. 91, no. 8, pp. 1247–1256, Aug. 2003.

[54] B. Sinopoli, C. Sharp, L. Schenato, S. Schaffert, and S. S. Sastry, "Distributed control applications within sensor networks," *Proc. of the IEEE*, vol. 91, no. 8, pp. 1235–1246, Aug. 2003.

[55] H. Qi, Y. Xu, and X. Wang, "Mobile-agent-based collaborative signal and information processing in sensor networks," *Proc. of the IEEE*, vol. 91, no. 8, pp. 1172–1183, Aug. 2003.

[56] A. H. Sayed and C. G. Lopes, "Distributed recursive least-squares strategies over adaptive networks," *Proc. Asilomar Conf. on Signals, Systems and Computers*, pp. 233–237, Oct. 2006.

[57] C. G. Lopes and A. H. Sayed, "Distributed adaptive incremental strategies: Formulation and performance analysis," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 3, pp. 584–587, May 2006.

[58] A. H. Sayed and C. G. Lopes, "Adaptive processing over distributed networks," *IEICE Trans. Fundamentals of Electorincs, Communications and Computer Sciences*, vol. E90-A, no. 8, pp. 1504–1510, Aug. 2007.

[59] C. G. Lopes and A. H. Sayed, "Incremental adaptive strategies over distributed networks," *IEEE Transactions on Signal Processing*, vol. 55, no. 8, pp. 4064–4077, Aug. 2007.

[60] S. Wehr, I. Kozintsev, R. Lienhart, and W. Kellermann, "Synchronization of acoustic sensors for distributed ad-hoc audio networks and its use for blind source separation," in *Proc. IEEE 6th Int. Symp. on Multimedia Software Eng.*, Dec. 2004, pp. 18–25.

[61] Y. Jia, Y. Luo, Y. Lin, and I. Kozintsev, "Distributed microphone arrays for digital home and office," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2006, vol. 5.

[62] S. Doclo, M. Moonen, T. Van den Bogaert, and J. Wouters, "Reduced-bandwidth and distributed MWF-based noise reduction algorithms for binaural hearing aids," *IEEE Trans. Audio, Speech and Language Processing*, vol. 17, no. 1, pp. 38–51, Jan. 2009.

[63] A. Bertrand and M. Moonen, "Distributed adaptive node-specific signal estimation in fully connected sensor networks – part I: Sequential node updating," *IEEE Transactions on Signal Processing*, vol. 58, no. 10, pp. 5277–5291, Oct. 2010.

[64] S. Markovich-Golan, S. Gannot, and I. Cohen, "A reduced bandwidth binaural MVDR beamformer," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Tel Aviv, Israel, Aug. 2010.

[65] T. C. Lawin-Ore and S. Doclo, "Analysis of rate constraints for MWF-based noise reduction in acoustic sensor networks," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011.

[66] A. Bertrand, "Applications and trends in wireless acoustic sensor networks: A signal processing perspective," *Proc. IEEE Symposium on Communications and Vehicular Technology (SCVT), (Ghent, Belgium)*, Nov. 2011.

[67] A. Markides, *Binaural Hearing Aids*, London, U.K.: Academic, 1977.

[68] S. Doclo, T. J. Klasen, T. Van den Bogaert, J. Wouters, and M. Moonen, "Theoretical analysis of binaural cue preservation using multi-channel wiener filtering and interaural transfer functions," *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, , no. 7, Sep. 2006.

[69] T. J. Klasen, T. Van den Bogaert, M. Moonen, and J. Wouters, "Binaural noise reduction algorithms for hearing aids that preserve interaural time delays cues," *IEEE Transactions on Signal Processing*, vol. 55, no. 4, pp. 1579–1585, Apr. 2007.

[70] S. Doclo, R. Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen, "Extension of the multi-channel wiener filter with localization cues for noise reduction in binaural hearing aids," *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, pp. 221–224, Sep. 2005.

[71] T. Van den Bogaert, S. Doclo, M. Moonen, and J. Wouters, "Binaural cue preservation for hearing aids using an interaural transfer function multichannel wiener filter," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 565–568, Apr. 2007.

[72] B. Cornelis, S. Doclo, T. Van den Bogaert, M. Moonen, and J. Wouters, "Analysis of localization cue preservation by multichannel wiener filter based binaural noise reduction in hearing aids," Aug. 2008.

[73] O. Roy and M. Vetterli, "Rate-constrained collaborative noise reduction for wireless hearing aids," *IEEE Transactions on Signal Processing*, vol. 57, no. 2, pp. 645–657, Feb. 2009.

[74] A. Bertrand and M. Moonen, "Distributed node-specific LCMV beamforming in wireless sensor networks," *IEEE Transactions on Signal Processing*, vol. 60, no. 1, pp. 233–246, Jan. 2012.

[75] J. P. Dmochowski, Z. Liu, and P. A. Chou, "Blind source separation in a distributed microphone meeting environment for improved teleconferencing," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2008, pp. 89–92.

[76] T. Ono, S. Miyage, N. Ono, and S. Sagayama, "Blind source separation with distributed microphone pairs using permutation correction by intra-pair TDOA clustering," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, 2010.

[77] J. Elson and R. Kay, "Wireless sensor networks: A new regime for time synchronization," *SIGCOMM Comput. Commun. Rev.*, vol. 33, no. 1, pp. 149–154, Jan. 2003.

[78] S. Ando and N. Ono, "A bayesian theory of cooperative calibration and synchronization in sensor networks," *Trans. Soc. Instru. Control Engineers (SICE)*, vol. E-S-1, pp. 21–26, 2005.

[79] N. Ono, H. Kohno, N. Ito, and S. Sagayama, "Blind alignment of asynchronously recorded signals for distributed microphone array," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, 2009, pp. 161–164.

[80] M. Pawig, G. Enzner, and P. Vary, "Adaptive sampling rate correction for acoustic echo control in voice-over-IP," *IEEE Transactions on Signal Processing*, vol. 58, no. 1, pp. 189–199, Jan. 2010.

[81] L. Erup, F.M. Gardner, and R.A. Harris, "Interpolation in digital modems, II- implementation and performance," *IEEE Trans. Commun.*, vol. 41, no. 6, pp. 998–1008, Jun. 1993.

[82] Z. Liu, "Sound source separation with distributed microphone arrays in the presence of clock synchronization errors," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Seattle, Washington, USA, Sep. 2008.

[83] Z. Liu, Z. Zhang, L. He, and P. Chou, "Energy-based sound source localization and gain normalization for ad hoc microphone arrays," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, 2007, vol. 2.

[84] M. Chen, Z. Liu, L. He, P. Chou, and Z. Zhang, "Energy-based position estimation of microphones and speakers for ad hoc microphone arrays," in *Proc. IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, 2007, pp. 22–25.

[85] K. Yao, R. E. Hudson, C. W. Reed, D. Chen, and F. Lorenzelli, "Blind beamforming on a randomly distributed sensor array system," *IEEE Journal on Selected Areas in Communications*, vol. 16, no. 8, pp. 1555–1567, Oct. 1998.

[86] K. C. Kerby and J. T. Bernard, "Sidelobe level and wideband behavior of arrays of random subarrays," *IEEE Trans. on Antennas and Propagation*, vol. 54, no. 8, pp. 2253–2262, Aug. 2006.

[87] M. F. A. Ahmed and S. A. Vorobyov, "Performance characteristics of collaborative beamforming for wireless sensor networks with Gaussian distributed sensor nodes," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 3249–3252, 2009.

[88] I. Kodrasi, T. Rohdenburg, and S. Doclo, "Microphone position optimization for planar superdirective beamforming," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011, pp. 109–112.

[89] A. Bertrand and M. Moonen, "Efficient sensor subset selection and link failure response for linear MMSE signal estimation in wireless sensor networks," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aalborg - Denmark, Aug. 2010, pp. 1092–1096.

[90] R.A. Horn and C.R. Johnson, *Matrix Analysis*, Cambridge University Press, 1985.

[91] W. Herbordt and W. Kellermann, "Analysis of blocking matrices for generalized sidelobe cancellers for non-stationary broadband signals," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2002, vol. 4, pp. IV–4187.

[92] A. Krueger, E. Warsitz, and R. Haeb-Umbach, "Speech enhancement with a GSC-like structure employing eigenvector-based transfer function ratios estimation," *IEEE Trans. Audio, Speech and Language Processing*, vol. 19, no. 1, pp. 206–219, Jan. 2011.

[93] C. Tseng and L.J. Griffiths, "A systematic procedure for implementing the blocking matrix in decomposed form," in *Proc. Asilomar Conf. on Signals, Systems and Computers*, Oct. 1988, vol. 2, pp. 808–812.

[94] S. Markovich-Golan, S. Gannot, and I. Cohen, "A weighted multichannel Wiener filter for multiple sources scenarios," in *the 27th convention of the Israeli Chapter of IEEE*, Eilat, Israel, Nov. 2012.

[95] S. Markovich-Golan, S. Gannot, and I. Cohen, "Subspace tracking of multiple sources and its application to speakers extraction," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, pp. 201–204, Mar. 2010.

[96] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed multiple constraints generalized sidelobe canceler for fully connected wireless acoustic sensor networks," *IEEE Trans. Audio, Speech and Language Processing*, 2012.

[97] S. Markovich-Golan, S. Gannot, and I. Cohen, "Distributed GSC beamforming using the relative transfer function," in *Proc. European Signal Processing Conf. (EUSIPCO)*, Aug. 2012, pp. 1274–1278.

[98] S. Markovich-Golan, S. Gannot, and I. Cohen, "Blind sampling rate offset estimation and compensation in wireless acoustic sensor networks with application to beamforming," in *Proc. Int. Workshop on Acoustic Echo and Noise Control (IWAENC)*, Aachen, Germany, Sep. 2012.

[99] S. Markovich-Golan, S. Gannot, and I. Cohen, "Performance analysis of a randomly spaced wireless microphone array," in *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, May 2011, pp. 121–124.

[100] S. Markovich-Golan, S. Gannot, and I. Cohen, "Performance of the SDW-MWF with randomly located microphones in a reverberant enclosure," *submitted for publication in IEEE Trans. Audio, Speech and Language Processing*, Aug. 2012.

[101] S. Markovich-Golan, S. Gannot, and I. Cohen, "A sparse blocking matrix for multiple constraints GSC beamformer," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, Mar. 2012.

[102] S. Markovich-Golan, S. Gannot, and I. Cohen, "Low-complexity addition or removal of sensors/constraints in LCMV beamformers," *IEEE Transactions on Signal Processing*, vol. 60, no. 3, pp. 1205–1214, Mar. 2012.

[103] E. Warsitz and R. Haeb-Umbach, "Acoustic filter-and-sum beamforming by adaptive principal component analysis," *Proc. IEEE Int. Conf. Acoustics, Speech, and Signal Processing (ICASSP)*, vol. 4, pp. 797–800, Mar. 2005.

[104] A. Bertrand and M. Moonen., "Distributed adaptive estimation of node-specific signals in wireless sensor networks with a tree topology," *IEEE Transactions on Signal Processing*, vol. 59, no. 5, pp. 2196–2210, May 2011.

[105] I. Cohen, "Relative transfer function identification using speech signals," *IEEE Trans. Speech and Audio Processing*, vol. 12, no. 5, pp. 451–459, Sep. 2004.

[106] B. Widrow and S.D. Stearns, "The LMS algorithm," in *Adaptive Signal Processing*, S. Haykin, Ed. Englewood Cliffs, NJ, Prentice-Hall, Inc., 1985.

[107] J.B. Allen and D.A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, no. 4, pp. 943–950, Apr. 1979.

[108] E.A.P Habets, "Room impulse response (RIR) generator," http://home.tiscali.nl/ehabets/rir_generator.html, Jul. 2006.

[109] J. J. Shynk, "Frequency-domain and multirate adaptive filtering," *IEEE Signal Processing Magazine*, vol. 9, no. 1, pp. 14–37, Jan. 1992.

[110] 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.

[111] M.R. Schroeder, "Statistical parameters of the frequency response curves of large rooms," *J. Audio Eng. Soc*, vol. 35, no. 5, pp. 299–306, 1987.

[112] J.D. Polack, "Playing billiards in the concert hall: The mathematical foundations of geometrical room acoustics," *Applied Acoustics*, vol. 38, no. 24, pp. 235–244, 1993.

[113] M.R. Schroeder, "Frequency correlation functions of frequency responses in rooms," *Journal of the Acoustical Society of America*, vol. 34, no. 12, pp. 1819–1823, 1962.

[114] H. Kuttruff, *Room Acoustics*, Spon, London, UK, 4th edition, 1999.

[115] J.M. Jot, L. Cerveau, and O. Warusfel, "Analysis and synthesis of room reverberation based on a statistical time-frequency model," in *Audio Engineering Society Convention 103*, Sep. 1997.

[116] F. Jacobsen and T. Roisin, "The cohrerence of reverberant sound fields," *Journal of the Acoustical Society of America*, vol. 108, no. 1, pp. 204–210, 2000.

[117] N. Dal-Degan and C. Prati, "Acoustic noise analysis and speech enhancement techniques for mobile radio applications," *Signal Processing*, vol. 18, pp. 43–56, 1988.

[118] H. L. Van Trees, *Optimum Array Processing: Part IV of Detection, Estimation, and Modulation Theory*, chapter Arrays and Spatial Filters, pp. 17–89, John Wiley & Sons, Inc., 2002.

[119] M.A. Woodbury, "Inventing modified matrices," *Statistical Research Group, Memo*, vol. Rep. no. 42, pp. Princeton University, Princeton N.J., 1950.

[120] A.S. Householder, "Unitary triangularization of a nonsymmetric matrix," *J. ACM*, vol. 5, pp. 339–342, Oct. 1958.

[121] G. H. Golub and C. F. van Loan, *Matrix Computations*, The Johns Hopkins University Press, 3rd ed., 1996.

עיבוד אותות דיבור מבוזר
עבור רשתות מיקרופונים

חיבור לשם קבלת התואר "דוקטור לפילוסופיה"

מאת:

שמוליק מרקוביץ'-גולן

הפקולטה להנדסה

הוגש לסנט של אוניברסיטת בר-אילן

רמת גן ניסן, תשע"ג

עבודה זו נעשתה בהדרכתם של:

פרופ׳ שרון גנות, הפקולטה להנדסה, אוניברסיטת בר-אילן

פרופ׳ ישראל כהן, הפקולטה להנדסת חשמל, הטכניון מט״ל

תודות

ברצוני להביע את תודתי והערכתי העמוקה למנחיי פרופ׳ שרון גנות ופרופ׳ ישראל כהן על הדרכתם והנחייתם המסורה. תודה על התמיכה המקצועית, על עידודכם לשלמות ועל הרבה עצות מועילות לכל אורך שלבי המחקר.

אני מודה לנשיא בר-אילן על התמיכה הכספית הנדיבה בהשתלמותי.

תודה לאמי אורית, לאבי המנוח יעקב, לשבתאי אהרון ולאלי ואורלי גולן. תודה מיוחדת לאהובתי לירן על תמיכתה האינסופית לאורך כל הדרך.

תוכן עניינים

תקציר

טכניקות שיפור דיבור, העושות שימוש במערכי מיקרופונים, לכדו את תשומת לבם של חוקרים רבים במשך שלושים השנים האחרונות, במיוחד עבור תרחישי תקשורת מסוג hands-free. בעיות אופיינות בתחום זה הן: הפחתת רעש, ביטול הד, מיצוי דובר וביטול הדהוד. אלגוריתמים מעצבי אלומה (Beamforming) מרחיבים את מימד הפתרונות ומציעים סינון מרחבי בנוסף לסינון זמן-תדר קלאסי. לרב, אותות הדיבור הנקלטים מזוהמים ע״י מקורות מפריעים, כגון דוברים מתחרים ומקורות רעש, ובנוסף מעוותים ע״י הסביבה המהדהדת. אף על פי שאלגוריתמים העושים שימוש במיקרופון בודד מציגים ביצועים מספקים בהפחתת רעש, הם אינם מתאימים דיים לביטול דוברים מפריעים, כיוון שהם חסרים את המידע המרחבי, או הגיוון הסטטיסטי המשמשים אלגוריתמים מרובי מיקרופונים. למרות היתרונות הברורים על פני מערכות מיקרופון-בודד, מערכי מיקרופונים הנהוגים כיום עדיין סובלים ממגבלות ביצועים.

האפרטורה הקטנה יחסית של מערכים קונבנציונליים מהווה גורם מגביל בביצועי אלגוריתמי עיבוד מרחבי, כיוון שהם דוגמים את שדה האקוסטי רק באופן מקומי, ולרב במרחק יחסית גדול ממקור/ות העניין. בתרחישים הללו, צפויים יחס אות לרעש (SNR) ויחס ההגעה ישירה להדהוד (DRR) נמוכים, הגורמים לפגיעה בביצועים של מערכי מיקרופונים רגילים 'מקובצים', בעלי אפרטורה קטנה. כתוצאה מכך, מערכי מיקרופונים מקובצים, קלאסיים, אינם מספקים פתרון שלם בתרחישים הבאים לדוגמא: 1) *מערכות חווית תקשורת קרובה*: מערכות ועידה מודרניות מנסות לא רק להעביר את תוכן השיחה באמינות, אלא גם לאפשר בנוסף חווית שיחה טבעית ואינטראקציה בין אנשים הרחוקים פיזית, כאילו הם נמצאים באותו חדר. שיחות טלפון רבות, במיוחד במצב hands-free, מופרעות ע״י רעש רקע, דוברים מפריעים והדהוד. 2) *בתים חכמים* הם שם למערכת אחודה לשליטה בכל מערכות הבית (תאורה, מיזוג ומכשירים אלקטרוניים כגון טלוויזיה, מערכת קול וכדומה). בהקשר זה, רשתות מיקרופונים חכמות הן מרכיב הכרחי לשליטה ובקרה במערכות אלו, וכן עבור תקשורת במקרי חירום. 3) *רשויות אכיפת חוק*, כגון משטרה וביטחון פנים, עושות שימוש במערכות ציתות ומעקב אקוסטיים במקומות ציבוריים כחלק משגרת פעילותן. לרב הפעילויות הנ״ל מתבצעות בסביבות עוינות. על המיקרופונים להיות פזורים בשטח נרחב בכדי להבטיח כיסוי מתאים של הדוברים הרצויים.

הגישה *הריכוזית* הישירה היא למקם חיישנים בסביבה רחבה ולהעביר את המידע הנקלט בהם *למרכז היתוך מידע* אשר בו מתבצע העיבוד. על אף האופטימליות, גישה פשוטה זו דורשת שידור של כמויות גדולות של מידע. נוסף על כך, הגישה הפשוטה הנ״ל רגישה לתקלה במרכז ההיתוך, אשר תשבית את המערכת כולה. חסרון נוסף, הנובע מהמבנה של הפתרון הריכוזי, הינו מסלולי התקשורת הארוכים בין חיישנים לבין מרכז ההיתוך, אשר עלולים לכלול מספר דילוגים (כאשר מרכז ההיתוך והחיישנים לא מסוגלים לתקשר ישירות), המתבטאים בקצב הסתגלות מואט לשינויים ברשת או בסביבה.

לאחרונה, גילויים טכנולוגיים בתחום הננו-טכנולוגיה, מערכות אלקטרו-מכניות ממוזערות, בשילוב עם שיפור יכולות התקשורת, הפכו את חזון רשתות החיישנים המבוזרות לבר-השגה. רשת חיישנים אלחוטית מכילה מספר צמתי (יח׳ רשת) המתקשרות ביניהן באופן אלחוטי. כל צומת מורכב מחיישן אחד או יותר, יחידת עיבוד

ויחידת תקשורת אלחוטית המאפשרת להם לשתף מידע. מטרת המערכת הינה לקלוט תופעה פיזיקלית מסויימת, לעבד אותה, ולהפיק תוצאה רצויה נדרשת. במערכות מבוססות מערכים קלאסיים, החישה והעיבוד של המידע הנרכש מרוכזים באתר בודד הקרוי *מרכז ההיתוך* . תופעה פיזיקלית בסביבה, יוצרת גל המתפשט במרחב. ככל שהחיישנים קרובים למקור התופעה, כן גדל יחס האות לרעש של האות הנקלט, המביא לשגיאות שערוך נמוכות ולאיכות גבוהה יותר במוצא תהליך עיבוד האות.

הרעיון של רשתות חיישנים אלחוטיות הוא לחלק את משאבי המערכת (חיישנים, יכולות עיבוד ומחוללי אותות) בין הצמתים ולאפשר מבנה סקאלבילי, עם כיסוי מלא של הסביבה, קל לפריסה, וחסין. הממשק האלחוטי מאפשר הרחבה של טווחי החישה מעבר לגבולות מערכות מרכזי היתוך קוויים. התפוצה של החיישנים בשטחים נרחבים מקנים כיסוי טוב יותר עם יחס אות לרעש, ויחס אות להדהוד גבוהים יותר והפחתת הרגישות למיקום האותות.

הפיתוחים הטכנולוגיים מציגים אתגרים חדשים: 1) **מגבלות תקשורת**: טופולוגיות רשת דינמיות בשילוב עם קישוריות מוגבלת וסוללה קצרת חיים דורשים מחקר נוסף בפיתוח שיטות תקשורת אלחוטית ופרוטוקולים יעילים, חסינים וסקלאביליים (לא יידון בעבודה זו). 2) **אלגוריתם**: אלגוריתמי עיבוד אותות קלאסיים מתוכננים בדר״כ עבור תרחישים ריכוזיים, בהם המידע מהחיישנים זמין לצורך עיבוד במרכז. אם כן, נחוצים אלגוריתמים מבוזרים חדשים, המשלבים קריטריוני אופטימיזציה חדשים, המתחשבים באילוצי הבעיה.

עבודה זו מתרכזת באתגר השני, בפרט, מטרתה לפתח אלגוריתמי עיבוד אותות מבוזרים. על האלגוריתמים המפותחים להיות בעלי סיבוכיות נמוכה ולצרוך רוחב פס תקשורת נמוך. כל זאת, בלא פשרה בנושא הביצועים (למרות שניתן להפעיל אלגוריתמים תת-אופטימליים). דרישה נוספת הנובעת מאופייה של הבעיה, היא הצורך באלגוריתמים חסינים, אשר אינם רגישים לתקלות בצמתים בודדים, ואשר מסוגלים להתמודד עם שינויים בקישוריות הרשת. באופן טבעי, מרבית האפליקציות דורשות אלגוריתמים המסתגלים לשינויי הסביבה, או המקורות הנצפים.

תקשורת (במיוחד ברשתות אלחוטיות) היא הפעולה הדורשת את מירב האנרגיה בצומת. גישת עיבוד אותות חלופית לעיבוד הריכוזי הינה עיבוד *מקומי* אשר בו כל צומת משתמש רק במדידות שלו, ללא קשר לצמתים אחרים, ובכך מייתר את הצורך בתקשורת. למרות שבגישה זו כמות התקשורת היא מינימלית, היא כמובן מאלצת מגבלות ביצועים, כיוון שהיא מנצלת רק חלק קטן מהמידע המצוי ברשת. מערכות רבות משתמשות בשיטות לדחיסת המידע בכדי להקטין את רוחב הפס הנדרש להעברתו למרכז ההיתוך. למרות פשטותן, שיטות אלו לא מתחשבות באלגוריתמי עיבוד האות המתבצעים במרכז ההיתוך. בנוסף, דחיסה עלולה להרוס את הקוהרנטיות בין אותות החיישנים, ולמנוע הפעלתם של אלגוריתמי מעצבי אלומה. אלגוריתמים מבוזרים מנסים להשיג את הביצועים של שיטת מרכז ההיתוך, ולהקטין באופן ניכר את רוחב הפס הדרוש לתקשורת. כל צומת מבצע עיבוד מקומי ומפיץ את התוצאות ברשת. ע״י שילוב איטרטיבי של החישוב המקומי עם המידע המופץ ברשת, אלגוריתמים מבוזרים מתכנסים למקביליהם הריכוזיים. שימו לב שהאיטרציות יכולות להתבצע על קובץ דגימות מוקלטות או באופן רקורסיבי בזמן.

במעצבי אלומה קלאסיים הגאומטריה של מערך המיקרופונים לרב נקבעת מראש כדי להתאים לבעיה הנתונה. עם זאת, במספר תרחישי רשתות חיישנים מבוזרות לא ניתן לשלוט על תצורת המערך. כיוון שביצועי מעצב האלומה תלויים בגאומטריה של המערך, ניתוח ביצועים סטנדרטי אינו מתאים לרשתות חיישנים אלחוטיות. ב-1964 הציע לו (Lo) לשלב מודלים סטטיסטיים במיקומי החיישנים. בהמשך, הוא פיתח מודל סטטיסטי לביצועי מעצב האלומה והגיע להבנה טובה יותר של תכונות המערך (כגון כיווניות, רוחב אלומה, וגובה אונות צד) עבור מעצב אלומה מסוג Delay and Sum. אולם, הניתוח היה מוגבל למקורות צרי סרט בסביבות לא מהדהדות, ועבור מעצבי אלומה פשוטים שאינם תלויים במידע.

תכנון אלגוריתמים מבוזרים עבור רשתות חיישנים אלחוטיות מצריך התחשבות בהיבטים נוספים. יכולת העיבוד המצויה בצמתים היא מוגבלת. בעקבות זאת, כמות החישובים לצורך בניית מעצב האלומה והפעלתו בצמתים היא מוגבלת. בנוסף, כיוון שבכל צומת קיים מקור שעון דוגם נפרד, סטיות תדר דגימה בין צמתים הן בלתי-נמנעות. הסטיות הנ״ל גורמות פוגעות בביצועי מעצבי האלומה.

התרומה של הדיסרטציה הזו היא משולשת. ראשית, גרסאות מבוזרות של מעצבי אלומה קלאסיים מפותחות. בפרט, אנו מציעים אלגוריתמי עיבוד דיבור מבוזרים בסביבה עם מספר דוברים המבוססים על קריטריון מזעור רמת הרעש במוצא תחת מערכת אילוצים לינאריים (LCMV). שנית, מפותחת מסגרת חדשה להערכת ביצועי מעצבי אלומה ברשתות מיקרופונים אלחוטיות. בעיבוד מרחבי קלאסי התצורה של המערך היא לרב מתוכננת לבעיה הנתונה, בניגוד לרשתות מיקרופונים אלחוטיות, אשר בהן יכולה התצורה להיות אקראית ודינאמית. הביצועים של מעצבי אלומה תלויי-מידע עם מיקרופונים הממוקמים באקראי מנותחים. הניתוח יכול לשמש כהנחיות תכנון לקביעת מספר הצמתים/מיקרופונים הנדרשים על מנת לעמוד בביצועים הנדרשים. שלישית, אנו מתייחסים לבעיית הפחתת הסיבוכיות בהפעלת מעצבי אלומה מבוזרים, ללא הקרבת הביצועים. האופי הדינאמי של המקורות, הסביבה וקישוריות הרשת, דורשים שינויים תכופים למעצב האלומה המופעל. שיטות יעילות לעדכון מעצב האלומה בהתאם לדינמיקה הנ״ל מפותחות.