Robust Multichannel Microphone Beamforming

Craig Anderson

A Thesis submitted to the Victoria University of Wellington in fulfilment of the requirements for the degree of Doctor of Philosophy

Victoria University of Wellington

2015

Abstract

In this thesis, a method for the design and implementation of a spatially robust multichannel microphone beamforming system is presented.

A set of spatial correlation functions are derived for 2D and 3D far-field/near-field scenarios based on von Mises(-Fisher), Gaussian, and uniform source location distributions. These correlation functions are used to design spatially robust beamformers and blocking beamformers (nullformers) designed to enhance or suppress a known source, where the target source location is not perfectly known — due to either an incorrect location estimate or movement of the target while the beamformers are active.

The spatially robust beam/null-formers form signal and interferer plus noise references which can be further processed via a blind source separation algorithm to remove mutual components — removing the interference and sensor noise from the signal path and vice versa. The noise reduction performance of the combined beamforming and blind source separation system approaches that of a perfect information MVDR beamformer under reverberant conditions.

It is demonstrated that the proposed algorithm can be implemented on low-power hardware with good performance on hardware similar to current mobile platforms using a four-element microphone array.

Acknowledgements

I would like to acknowledge my supervisors Mark Poletti and Paul Teal for their guidance over the last few years during this PhD project and the preceding Masters' thesis; Alan Murray at Tait Communications for the useful industry input on this project (and introducing me to the many fine craft beer establishments in Christchurch); Walter Kellermann and Stefan Meier at Friedrich-Alexander Universität Erlangen-Nürnberg for their collaboration on the blind source related parts of this thesis, and for hosting me at Erlangen for ten weeks along with the rest of the LMS team; to friends, family, and colleges at Victoria and Callaghan Innovation for their support over the last few years; and to Ayla for not getting too upset/jealous over my trip to Germany (and the various other places I visited while over there).

I would also like to acknowledge the Ministry of Business Innovation and Employment for providing the PhD scholarship which funded this research, and the Royal Society of New Zealand for funding my travel to Germany as part of Walter Kellermann's Julius von Haast Fellowship award.

Contents

Chapter 1

Introduction

1.1 Motivation

This thesis is focussed on developing beamforming solutions for speech enhancement in noisy environments, with a particular focus on robustness to spatial errors (imperfect knowledge of where the source is relative to the microphone array) and have a computationally efficient implementation. The objective is to design a system which works reliably under noisy conditions without relying on user intervention or perfect knowledge of where the user is relative to the array, and is tolerant of array imperfections (microphone mismatching, geometry errors, and general noise). In addition, one of the objectives is to consider more realistic near-field and scatterer-based modelling of sound propagation which has largely been overlooked in the literature. The outputs of this thesis are intended to be used in the field of public safety, in particular for use in mobile devices. In this application a number of constraints are required, including a restriction of the number of microphones, geometry, size of the array, working in high background noise levels, and assuming limited (or no) user interaction with the mobile device to tweak parameters for example. The mobile device nature also implies approximate (but not perfect) knowledge of where the desired signal (the user) is in relation to the device, introducing the requirement of spatial robustness to beamforming algorithms.

1.2 Background

This thesis builds on the work previously done in [Anderson, 2012] in which a number of microphone beamforming methods were analysed in the context

of near-field beamforming for speech enhancement. Previously a variety of microphone beamforming methods such as differential array processing, least squares designs and adaptive designs were investigated in the context of speech enhancement in extreme noise environments. It was found that adaptive processing provided the best overall performance, however the system developed was not particularly robust to intrinsic sensor noise or movement.

1.2.1 Speech Enhancement

Previous approaches to speech enhancement in the literature focus on a number of techniques including multi-channel beamforming [Capon, 1969; Frost, 1972; Flanagan et al., 1985] (Section 2.3, Chapters 4 and 5), in which multiple microphone signals are processed to enhance speech by spatially suppressing background interference; post-processing methods such as spectral subtraction [Boll, 1979] and Wiener filtering [Jeub and Vary, 2010; Van Trees, 2004] (Section 2.3.5, Chapter 3), in which statistics of speech signals and known interference/noise signals are used to filter the microphone signals; blind source separation [Jolliffe, 2002; Hyvärinen et al., 2004; Kellermann et al., 2006] (Section 2.4, Chapter 6, Section 7.2) in which information theoretical approaches are used to separate mixtures of different sources (multiple talkers, speech in background noise, etc.); and more computationally expensive techniques such as speech dictionary training-based methods such as non-negative matrix factorisation [Wilson et al., 2008; Weninger et al., 2012], in which speech is enhanced through feature comparison/extraction methods.

1.3 Proposed Algorithm

In this thesis, a low-complexity robust adaptive beamforming algorithm will be developed using spatial correlation models designed to account for uncertainty in the expected location of the desired source. Two beamformers will be designed, one of which imperfectly enhances the desired signal, and the other imperfectly estimates the interference/noise. The outputs of the two beamformers will be post-processed using a blind source separation algorithm in order to remove mutual components from each channel: desired speech signal in the interference/noise channel, and interference/noise in the desired speech channel. The intent of the BSS system is to emulate a generalised sidelobe canceller-type design without the signal leakage problems inherent in that particular design.

1.4 Thesis Contents

The second chapter will introduce important concepts in spatial beamforming, the main focus of the thesis. An overview of current beamforming and blind source separation techniques is presented.

Chapter 3 will introduce a simple technique for obtaining a Wiener filter in real-time using the concept of dual robust beamformers to estimate signal and interference/noise statistics. This chapter develops the concept of the spatially robust null-steering beamformer (nullformer) to estimate interference.

Chapter 4 develops new novel spatial correlation models for far-field and near-field multi-channel beamformers using the von Mises(-Fisher), radial Gaussian, and uniform distributions. The key contributions to the literature

are the development of a set of cylindrical and spherical Bessel function expressions for far and near-field spatial correlation functions, which can be used to develop spatially robust beamformers, and of particular interest in this thesis, nullformers for interference estimation.

Chapter 5 develops models for designing beamformers for sources near or located on a solid sphere, and compares the theoretical performance of a scatterer-based design with traditional free-field designs. The intention is to more realistically model the effect of the human head on wave propagation, rather than assuming free-field propagation, and investigate whether there are any advantages in doing so. This has been a neglected aspect of prior work.

Chapter 6 combines the spatially robust fixed beamformer design with a two-channel blind source separation post-processor to compensate for imperfections in assumed knowledge used to design the beamformers. The intention is to use a well known blind source separation algorithm to emulate a traditional adaptive noise canceller design without the usual issues which arise with imperfect beamformer design.

Chapter 7 is focussed on the real-time implementation of the algorithms. The first of which is a graphics processing unit (GPU) accelerated blind source separation algorithm; and the second is a complete implementation of the fixed beamformer plus BSS post-processor solution introduced in Chapter 6. The results in this chapter demonstrate that the algorithms can be feasibly implemented on current devices.

Chapter 8 outlines the conclusions and highlights potential future research to extend this project.

1.5 Publications

The majority of this thesis has been submitted for publication, with the exception of the preliminary theory and portions of Chapter 7 focussed on implementations and real-world algorithm testing. The content is largely unchanged from the submitted/published versions aside from some repetitive information removal, corrections, and notation modifications to standardise throughout the thesis.

- Chapter 3 has been presented at the Statistical Signal Processing (SSP) conference held in 2014, as the paper "Multichannel Wiener Filter Estimation using Source Location Knowledge for Speech Enhancement" [Anderson et al., 2014b].

- Chapter 4, Section 1 has been published in the IEEE/ACM Transactions on Audio, Speech and Language Processing journal as the article: "Spatially Robust Far-field Beamforming using the von Mises(-Fisher) Distribution" [Anderson et al., 2015b]. Chapter 4, Section 2 has been submitted as the article: "Spatial Correlation of Spherically Symmetric Near-field Source Distributions", and has been accepted for publication in the IEEE/ACM Transactions on Audio, Speech and Language Processing journal.

- Chapter 6 was presented at the International Conference on Acoustics, Speech, and Signal Processing (ICASSP) conference held in 2015, as the paper "TRINICON-BSS System Incorporating Robust Dual Beamformers for Noise Reduction" [Anderson et al., 2015a].

- Chapter 7, section 1 was presented at the 4th Joint Workshop on

Hands-free Speech Communication and Microphone Arrays (HSCMA) conference held in 2014, as the paper "A GPU-Accelerated Real-time Implementation of TRINICON-BSS for Multiple Separation Units" [Anderson et al., 2014a].

All of the submitted articles/conference papers had input from my supervisors Paul Teal and Mark Poletti, with additional contributions from Stefan Meier and Walter Kellermann on the two TRINICON related papers.

Chapter 2

Preliminaries

2.1 Introduction

This chapter introduces the wave propagation models used to design beamformers for near and far-field scenarios. An overview of existing optimal beamforming algorithms is presented, and a brief introduction to blind source separation is provided.

2.2 Wave Propagation

2.2.1 Wave/Helmholtz Equation

Beamformer design relies on modelling the wave propagation between sound sources and a set of microphones in an array (or the relative propagation within the array). This is governed by the wave equation [Morse et al., 1948, p 294], which can be expressed as

$$\bigtriangledown^2 \psi - \frac{1}{c_0^2}\frac{\partial^2 \psi}{\partial t^2} = 0 \tag{2.1}$$

where ψ is the wave function and is a function of space and time, c_0 is the speed of the medium, which throughout this thesis was assumed to be $c_0 = 343.0\,\mathrm{ms}^{-1}$ — the speed of sound at 20°C, and the Laplace operator $\bigtriangledown^2$ defined in Cartesian coordinates as

$$\bigtriangledown^2 = \frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2} + \frac{\partial^2}{\partial z^2} \tag{2.2}$$

or in spherical coordinates as

$$\bigtriangledown^2 = \frac{1}{r^2}\frac{\partial}{\partial r}\left(r^2 \frac{\partial}{\partial r}\right) + \frac{1}{r^2 \sin\theta}\frac{\partial}{\partial \theta}\left(\sin\theta \frac{\partial}{\partial \theta}\right) + \frac{1}{r^2 \sin^2\theta}\frac{\partial^2}{\partial \phi^2} \tag{2.3}$$

Additionally, the Laplace operator can be defined in cylindrical, prolate spheroid, and many more coordinate systems. In this thesis the spherical coordinate system is the predominant system used.

Solutions to the wave equation can be found by assuming that the wave function can be decomposed into spatial and temporal components, and using the technique of separation of variables to solve, i.e.,

$$\psi = \psi_{\mathrm{sp.}}(x, y, z)\psi_{\mathrm{t}}(t) \tag{2.4}$$

or in spherical coordinates

$$\psi = \psi_{\text{sp.}}(r, \theta, \phi)\psi_{\text{t}}(t) \tag{2.5}$$

Inserting either of these definitions into (2.1) and dividing by $\psi_{sp.}\psi_t$ results in

$$\frac{1}{\psi_{\text{sp.}}} \bigtriangledown^2 \psi_{\text{sp.}} = \frac{1}{c_0^2 \psi_{\text{t}}} \frac{d^2\psi_{\text{t}}}{dt^2} \tag{2.6}$$

The equality in (2.6) is possible if both sides of are equal to some constant $-\beta_0$.

The temporal component can now be expressed as

$$\frac{d^2\psi_{\text{t}}}{dt^2} + \beta_0 c_0^2 \psi_{\text{t}} = 0 \tag{2.7}$$

A reasonable expectation for the time dependent component of the ψ function is that it is harmonic, thus we expect solutions of the form

$$\psi_{\text{t}}(t) \propto \sin(\omega t), \cos(\omega t), e^{i\omega t}, e^{-i\omega t} \tag{2.8}$$

Inserting one of the trial solutions into (2.7) gives

$$-\omega^2 \psi_{\text{t}} + \beta_0 c_0^2 \psi_{\text{t}} = 0 \tag{2.9}$$

(2.9) is satisfied if

$$\beta_0 = \frac{\omega^2}{c_0^2} \tag{2.10}$$

which is equal to the wavenumber k squared.

In spatial beamforming, the spatial component of the wave equation is of most interest. In (2.10) the solution for the constant β_0 was demonstrated to be the wavenumber squared. Inserting the constant in place of the time dependent component in (2.1) leads to the Helmholtz equation — the time

independent wave equation which will form the basis of the beamforming methods throughout this thesis.

The homogeneous Helmholtz equation is given as [Li and Duraiswami, 2007, (2)]

$$\bigtriangledown^2 \psi + k^2 \psi = 0 \tag{2.11}$$

2.2.2 Plane Wave Solution

The simplest, and most commonly used propagation model is the plane wave. In this model, wave propagation occurs, without attenuation, along some axis with a planar wave-front perpendicular to the travel direction.

Taking the x-axis as the propagation axis, the Helmholtz equation in one dimension becomes

$$\frac{d^2\psi_{\mathrm{x}}}{dx^2} + k^2 \psi_{\mathrm{x}} = 0 \tag{2.12}$$

which has solutions in terms of complex exponential functions/trigonometric functions

$$\psi_{\mathrm{x}}(x) \propto e^{ikx}, e^{-ikx}, \sin(kx), \cos(kx) \tag{2.13}$$

2.2.3 Spherically Symmetric Point Source

The plane-wave model is a simplification valid when measuring the sound field at a large distance from the source, commonly referred to as the far-field. For sources close to the measurement point, a model which includes attenuation is desirable (referred to as a near-field model). The near/far-field transition point is usually defined as occurring at some multiple of the wavelength, defined in [Mailloux, 2005, (1.47)] as $r_{\mathrm{ff}} = 2d^2/\lambda$, where d is the length of a linear array or the diameter of a circular/spherical array, and λ is the

wavelength. Below this point, the wave-fronts exhibit significant spherical curvature and attenuation across the array and can therefore no longer be accurately modelled as plane-waves.

A basic near-field model of wave propagation can be developed by considering a spherically symmetrical source, i.e., there are no angular variations in the wavefronts, and allowing the propagating wave to decay with increasing distance from the source (as seen in Figure 2.1).

The spherical coordinate definition of the Helmholtz equation can be expressed as

$$\nabla_r^2 \, \psi_{\mathrm{r}} + k^2 \, \psi_{\mathrm{r}} = 0 \tag{2.14}$$

It can be demonstrated that the trial solution

$$\psi_{\mathrm{r}}(r) = \frac{e^{ikr}}{r} \tag{2.15}$$

is a solution of the Helmholtz equation. This trial solution provides a simple description of a spherically symmetric point source, which will be an important result for the design of near-field beamformers in subsequent chapters.

Expanding (2.14) gives,

$$\frac{d^2\psi_{\mathrm{r}}}{dr^2} + \frac{2}{r}\frac{d\psi_{\mathrm{r}}}{dr} + k^2 \, \psi_{\mathrm{r}} = 0 \tag{2.16}$$

The first derivative of the ψ_{r} function is

$$\frac{d\psi_{\mathrm{r}}}{dr} = -\frac{e^{ikr}}{r^2} + ik\frac{e^{ikr}}{r} \tag{2.17}$$

$$= \left(-\frac{1}{r} + ik\right)\psi_{\mathrm{r}} \tag{2.18}$$

The second derivative of the ψ_{r} function is given as

$$\frac{d^2\psi_{\mathrm{r}}}{dr^2} = 2\frac{e^{ikr}}{r^3} - 2ik\frac{e^{ikr}}{r^2} - k^2\frac{e^{ikr}}{r} \tag{2.19}$$

$$= \left(\frac{2}{r^2} - \frac{2ik}{r} - k^2\right)\psi_{\mathrm{r}} \tag{2.20}$$

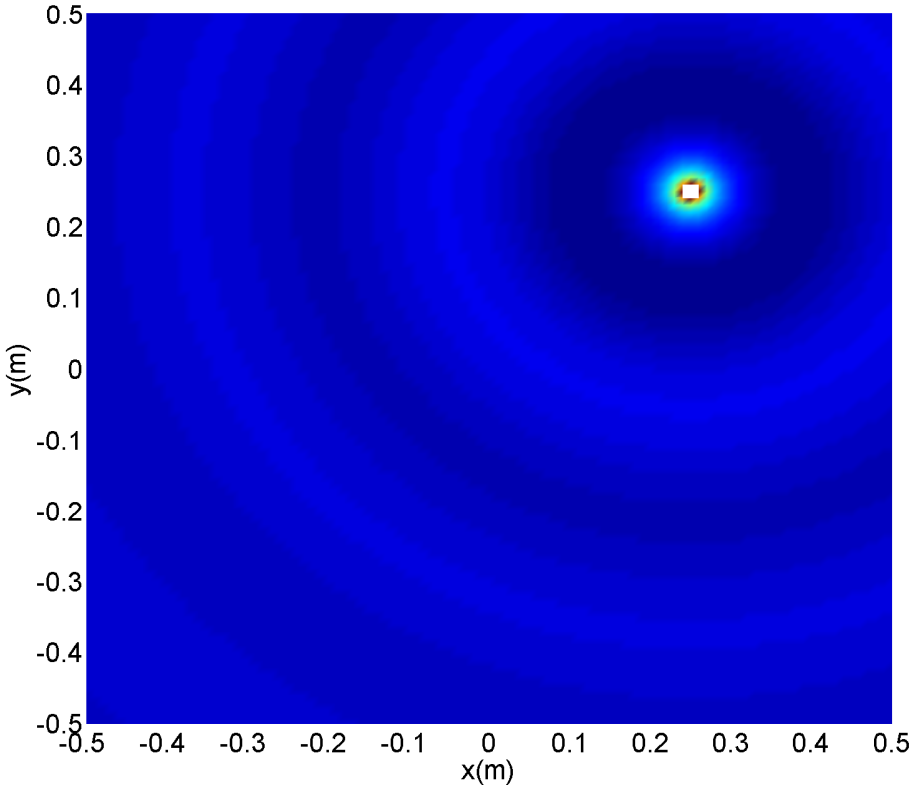

Figure 2.1: Near-field point source propagation. The wavefronts expand outwards from the source origin symmetrically.

Inserting (2.18) and (2.20) into the left hand side of (2.16) gives

$$\left(\frac{2}{r^2} - \frac{2ik}{r} - k^2\right)\psi_{\mathrm{r}} + \left(-\frac{2}{r^2} + \frac{2ik}{r}\right)\psi_{\mathrm{r}} + k^2\,\psi_{\mathrm{r}} = 0 \qquad (2.21)$$

from which it can be seen that the terms cancel, therefore the simple point source equation is a valid solution of the Helmholtz equation.

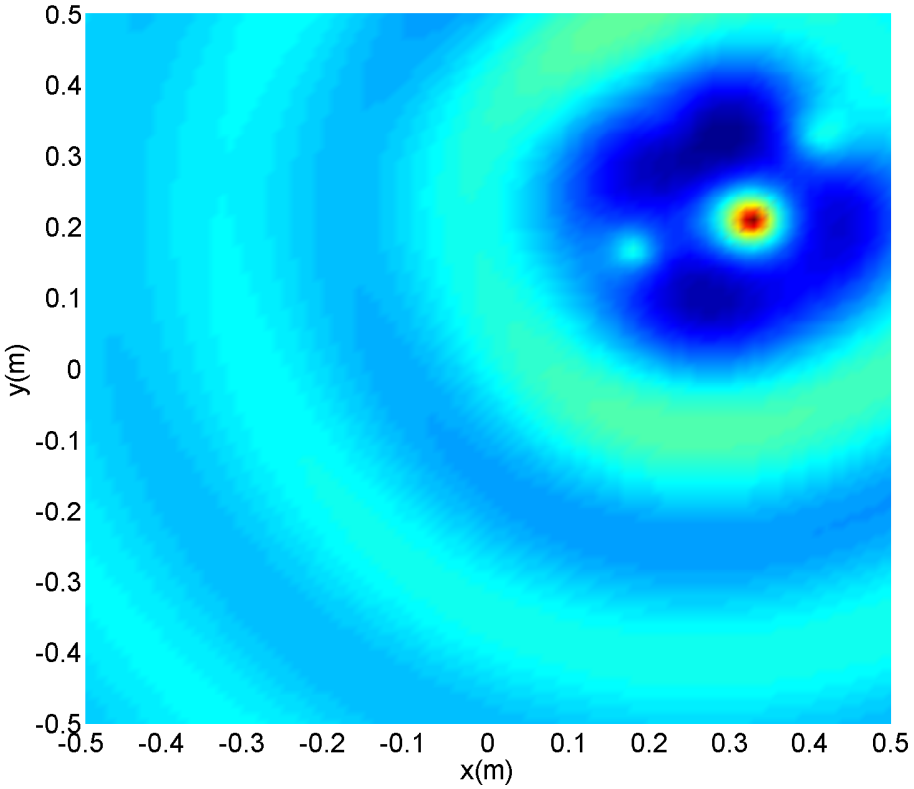

Figure 2.2: A somewhat more complicated source.

2.2.4 General Solution

A general solution in spherical coordinates can be obtained by considering angular components as well as the radial component, this allows for solutions for more complex wave propagation such as the example displayed in Figure 2.2. Assuming the wave function is separable into radial and angular components, it can be expressed as

$$\psi(r,\theta,\phi) = \psi_{\mathrm{r}}(r)\psi_{\theta}(\theta)\psi_{\phi}(\phi) \tag{2.22}$$

Inserting into the Helmholtz equation, and using the spherical coordinate definition of the Laplacian gives

$$\frac{\psi_\theta\psi_\phi}{r^2}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi_\mathrm{r}}{\partial r}\right)+\frac{\psi_\mathrm{r}\psi_\phi}{r^2\sin\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial\psi_\theta}{\partial\theta}\right)+\frac{\psi_\mathrm{r}\psi_\theta}{r^2\sin^2\theta}\frac{\partial^2\psi_\phi}{\partial\phi^2}+k^2\psi_\mathrm{r}\psi_\theta\psi_\phi=0 \tag{2.23}$$

Dividing (2.23) by $\psi_\mathrm{r}\psi_\theta\psi_\phi$ and multiplying by r^2 gives,

$$\frac{1}{\psi_\mathrm{r}}\frac{\partial}{\partial r}\left(r^2\frac{\partial\psi_\mathrm{r}}{\partial r}\right)+k^2r^2+\frac{1}{\sin\theta\psi_\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial\psi_\theta}{\partial\theta}\right)+\frac{1}{\sin^2\theta\psi_\phi}\frac{\partial^2\psi_\phi}{\partial\phi^2}=0 \tag{2.24}$$

in which the radial and angular parts are now separated. This implies that the radial and angular parts can be expressed as being equal to some constant β_1,

$$\frac{1}{\psi_\mathrm{r}}\frac{d}{dr}\left(r^2\frac{d\psi_\mathrm{r}}{dr}\right)+k^2r^2=-\frac{1}{\sin\theta\psi_\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial\psi_\theta}{\partial\theta}\right)+\frac{1}{\sin^2\theta\psi_\phi}\frac{\partial^2\psi_\phi}{\partial\phi^2}=\beta_1 \tag{2.25}$$

The angular component can be separated into θ and ϕ components following a similar treatment.

$$\beta_1+\frac{1}{\sin\theta\psi_\theta}\frac{\partial}{\partial\theta}\left(\sin\theta\,\frac{\partial\psi_\theta}{\partial\theta}\right)+\frac{1}{\sin^2\theta\psi_\phi}\frac{\partial^2\psi_\phi}{\partial\phi^2}=0 \tag{2.26}$$

Multiplying by $\sin^2\theta$ separates the θ and ϕ components as follows,

$$\beta_1\sin^2\theta+\frac{\sin\theta}{\psi_\theta}\frac{d}{d\theta}\left(\sin\theta\,\frac{d\psi_\theta}{d\theta}\right)+\frac{1}{\psi_\phi}\frac{d^2\psi_\phi}{d\phi^2}=0 \tag{2.27}$$

The two independent components are now equal to some constant $-\beta_2$

$$\beta_1\sin^2\theta+\frac{\sin\theta}{\psi_\theta}\frac{d}{d\theta}\left(\sin\theta\,\frac{d\psi_\theta}{d\theta}\right)=-\frac{1}{\psi_\phi}\frac{d^2\psi_\phi}{d\phi^2}=-\beta_2 \tag{2.28}$$

It can now be seen that there are now separate differential equations for

each separable function:

$$\frac{d}{dr}\left(r^2\frac{d\psi_{\rm r}}{dr}\right)+\left(k^2r^2-\beta_1\right)\psi_{\rm r}=0 \tag{2.29}$$

$$\sin\theta\,\frac{d}{d\theta}\left(\sin\theta\,\frac{d}{d\theta}\psi_\theta\right)+\left(\beta_1\sin^2\theta-\beta_2\right)\psi_\theta=0 \tag{2.30}$$

$$\frac{d^2\psi_\phi}{d\phi^2}-\beta_2\psi_\phi=0 \tag{2.31}$$

Angular Solutions

Equation (2.31) has solutions described by the complex exponential functions

$$\psi_\phi=e^{i\sqrt{\beta_2}\phi},e^{-i\sqrt{\beta_2}\phi} \tag{2.32}$$

Boundary conditions require that $\psi_\phi(0)=\psi_\phi(2\pi)$, which is satisfied if $\sqrt{\beta_2}$ is equal to some integer m.

The ψ_θ component can be obtained by defining the following

$$x=\cos\theta \tag{2.33}$$

$$\frac{dx}{d\theta}=-\sin\theta \tag{2.34}$$

and finding a solution in terms of a power series in x [Morse and Ingard, 1968, (p333-334)].

Equation (2.30) can be re-expressed after some manipulation as

$$(1-x^2)\frac{d^2\psi_\theta}{dx^2}-2x\frac{d\psi_\theta}{dx}+\left(\beta_1-\frac{m^2}{(1-x^2)}\right)\psi_\theta=0 \tag{2.35}$$

which has solutions in terms of associated Legendre functions [Morse and Ingard, 1968, (p 333-334)], requiring $\beta_1=n(n+1)$ (where n is an integer) to prevent divergence for $\cos\theta=\pm1$:

$$\psi_\theta\propto P_n^m(\cos\theta) \tag{2.36}$$

Combining the θ and ϕ angular functions results in the spherical harmonic functions,

$$Y_n^m(\theta,\phi) \equiv \alpha_n^m P_n^m(\cos\theta) e^{im\phi} = \alpha_n^m \psi_\theta(\theta)\psi_\phi(\phi) \tag{2.37}$$

where

$$\alpha_n^m = \sqrt{\frac{(2n+1)(n-m)!}{4\pi(n+m)!}} \tag{2.38}$$

is a normalisation constant to ensure orthonormality:

$$\int_\Omega Y_n^m(\theta,\phi) Y_n^{m*}(\theta,\phi)\, \sin\theta\, d\theta\, d\phi = 1 \tag{2.39}$$

where Ω indicates a surface integral.

Radial Solutions

The radial equation (2.29) can be expressed as

$$r^2 \frac{d^2\psi_\mathrm{r}}{dr^2} + 2r\frac{d\psi_\mathrm{r}}{dr} + \left(k^2 r^2 - n(n+1)\right)\psi_\mathrm{r} = 0 \tag{2.40}$$

which has solutions in terms of spherical Bessel/Hankel functions, found by considering a power series solution in r [Morse and Ingard, 1968, (p 6, 336)].

The radial component can be expressed as

$$\psi_\mathrm{r}(r) \propto j_n(kr), y_n(kr), h_n(kr), h_n^*(kr) \tag{2.41}$$

where j_n is the spherical Bessel function of the first kind, y_n is the spherical Bessel function of the second kind, and h_n is the spherical Hankel function:

$$h_n(kr) = j_n(kr) + iy_n(kr) \tag{2.42}$$

General Result

Since a linear combination of solutions to the Helmholtz equation is itself also a solution of the Helmholtz equation, a general solution can be expressed in terms of weighted Bessel and spherical harmonic functions. The general solution for the spatial component of the wave equation can therefore be expressed as

$$\psi(r,\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m j_n(kr) Y_n^m(\theta,\phi) \tag{2.43}$$

$$\psi(r,\theta,\phi) = \sum_{n=0}^{\infty} \sum_{m=-n}^{n} A_n^m h_n(kr) Y_n^m(\theta,\phi) \tag{2.44}$$

where A_n^m denotes the arbitrary combined Bessel and spherical harmonic coefficients, and the choice of j_n or h_n depends on the propagation scenario to be considered. In later chapters these will be described in more detail. The y_n and h_n^* based solutions (from (2.41)) are also valid, but not used in this thesis. The y_n based solution diverges at the origin; and the h_n^* solutions represent travelling waves in the opposite direction to the h_n solutions.

2.3 Beamforming

2.3.1 Delay/Phase and Sum Beamformer

The output of a microphone array at a particular wavenumber/frequency k and time-index i (in the short time-frequency domain) can be expressed as

$$\mathbf{x}(i,k) = s(i,k)\boldsymbol{\psi}(i,k) + \mathbf{n}(i,k) \tag{2.45}$$

where $\mathbf{x}$ is a vector of received microphone signals, s denotes the desired signal amplitude, $\boldsymbol{\psi}$ the source to microphone transfer function vector — the

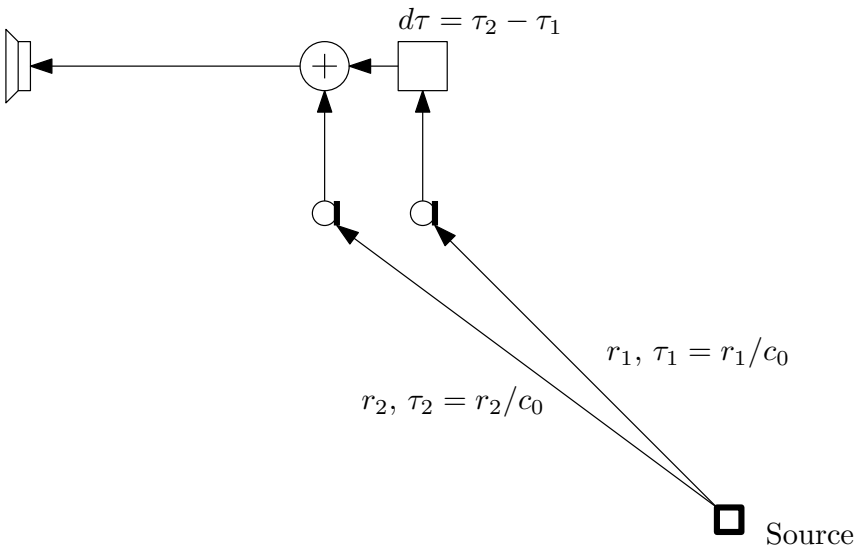

Figure 2.3: Simple delay and sum beamformer. Here waves travelling from the source take two different lengths of time to reach each of the sensors in the array. The delay element above the closest element time/phase-aligns the signals received.

set of source to microphone wave propagation functions, and $\mathbf{n}$, the sensor noise, which is assumed to be uncorrelated between each sensor.

The delay/phase and sum beamformer [Benesty et al., 2008; Flanagan et al., 1985; Flanagan et al., 1991a; Flanagan et al., 1991b; Yardibi et al., 2010] can be described as a set of weights $\mathbf{w}$ which, when applied to the signals received at the microphone array, compensate for the relative delay/phase differences between the microphones due to some desired source. From Figure 2.3 it can be seen that the different path lengths between the source and the two microphones will lead to a difference in the time of arrival of the

wavefront as it propagates from the source. The signal from the source can be time-aligned by introducing a delay at the closest element equal to the time difference of arrival between the two elements. Using the spatial-frequency analysis, this is equivalent to a phase difference between the signals received at the microphones. The signals can be phase aligned by multiplying the output of the first microphone by some complex weight corresponding to this phase difference.

The simple 2-microphone beamformer example, where the objective is to find the weight w_1 to double the amplitude (sum the time/phase aligned desired signal) by time/phase aligning the microphone signals, can be solved as

$$\begin{aligned} w_1^* x_1 + x_2 &= 2x_2 \\ w_1^* \left[s\, \psi_{s,1} \right] + \left[s\, \psi_{s,2} \right] &= 2 \left[s\, \psi_{s,2} \right] \\ w_1 &= \left(\frac{\psi_{s,2}}{\psi_{s,1}} \right)^* \end{aligned}$$

where w_1 now compensates for both the phase difference and amplitude difference between the microphones in the 2-element example (for a single frequency). The end result of this time/phase alignment is to electronically steer the response of the array towards the desired source. This can be trivially extended to more than two microphones by choosing one as the reference microphone, and designing the weights such that all of the relative phases between the microphones and the reference microphone cancel.

2.3.2 Interference Suppression

Suppose now, as an example, there are N undesired sources active near the microphone array. The signals received at the microphone array can now be

expressed as

$$\mathbf{x} = s\boldsymbol{\psi}_{\mathrm{s}} + \sum_{n=1}^{N} v_n \boldsymbol{\psi}_{\mathrm{v}_n} + \mathbf{n} \tag{2.46}$$

where v_n is an interferer signal, and $\boldsymbol{\psi}_{\mathrm{v}_n}$ the interferer to microphone array propagation vector. The time/frequency indexing has been dropped for clarity. Unlike sensor noise, interference from one or more other sound sources can be (and usually is) correlated across the microphones in the array. For the interferer canceller case, the objective of the beamformer would be to simultaneously direct the response towards the desired signal and block the interferer. Mathematically this can be expressed as trying to solve the simultaneous equations

$$\begin{aligned} w_1\psi^*_{\mathrm{s},1} + w_2\psi^*_{\mathrm{s},2} + ... + w_M\psi^*_{\mathrm{s},M} &= 1 \\ w_1\psi^*_{\mathrm{v}_1,1} + w_2\psi^*_{\mathrm{v}_1,2} + ... + w_M\psi^*_{\mathrm{v}_1,M} &= 0 \\ \vdots \qquad\qquad &= 0 \\ w_1\psi^*_{\mathrm{v}_N,1} + w_2\psi^*_{\mathrm{v}_N,2} + ... + w_M\psi^*_{\mathrm{v}_N,M} &= 0 \end{aligned} \tag{2.47}$$

This can be expressed in matrix form as

$$\boldsymbol{\Psi}^H \mathbf{w} = \mathbf{c} \tag{2.48}$$

where $\mathbf{w}$ is the vector containing the beamformer weights $(w_1, w_2, ..., w_M)$ to solve for, $\boldsymbol{\Psi}$ the matrix containing the transfer functions for each of the source to microphone pairs (the ψ functions in (2.47)), $\mathbf{c}$ the constraint vector (the right-hand side of (2.47)), and the ^{H} symbol denotes the conjugate transpose (Hermitian) operator.

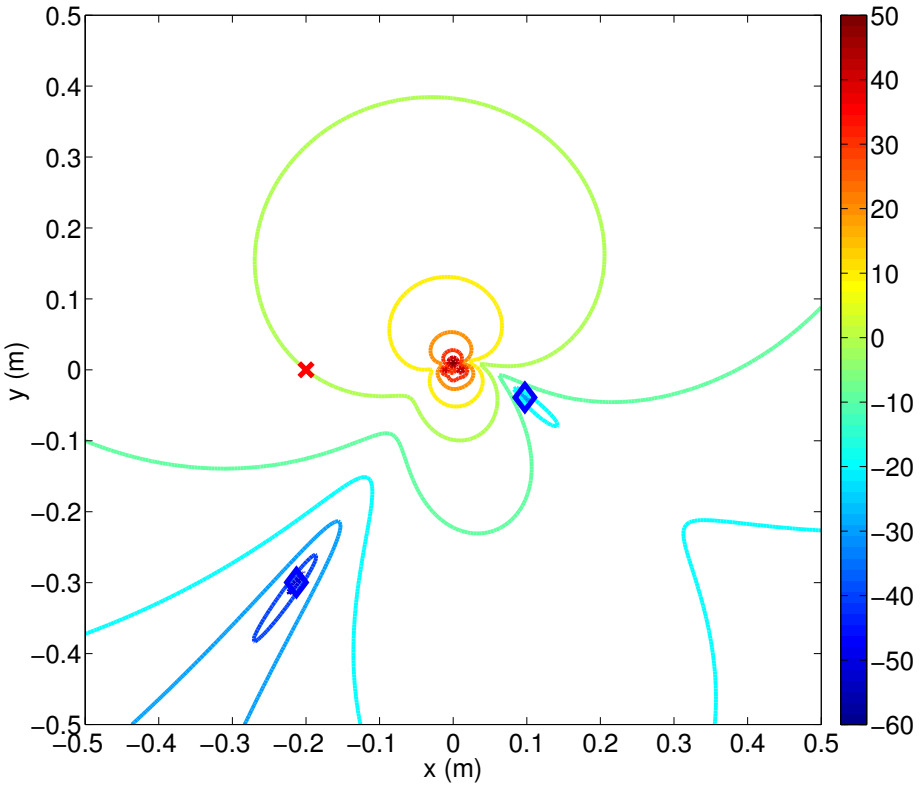

Figure 2.4: Near-field beampattern (in dB) designed using (2.50) for a single target and two interferers using a 4-element circular microphone array.

The beamformer weights can be solved for using least squares methods [Anderson, 2012] as

$$\left(\mathbf{\Psi}\mathbf{\Psi}^H\right)\mathbf{w} = \mathbf{\Psi}\mathbf{c} \tag{2.49}$$

$$\mathbf{w} = \left(\mathbf{\Psi}\mathbf{\Psi}^H\right)^{-1}\mathbf{\Psi}\mathbf{c} \tag{2.50}$$

In Figure 2.4, a simple demonstration of the least squares method is shown. The beamformer has been designed to pass the desired source located at (-0.20 m, 0 m) undistorted and reject the two interferers at (-0.21 m, -0.30 m)

and (0.10 m, -0.04 m) — corresponding to the constraint vector

$$\mathbf{c}^T = \begin{bmatrix} 1 & 0 & 0 \end{bmatrix} \tag{2.51}$$

As it will be seen in the next subsection, this simple interference blocking beamformer is closely related to both the Minimum Variance Distortionless Reponse (MVDR) and the more general Linearly Constrained Minimum Variance (LCMV) beamformers up to a simple scaling factor.

2.3.3 MVDR/LCMV Beamforming

An optimal beamformer problem can be specified as trying to minimise the interference and noise while simultaneously ensuring the desired signal is received undistorted. This is the known as the Minimum Variance Distortionless Response (MVDR)/Capon beamformer [Capon, 1969; Haykin, 1991; Moonen, 1993; Lorenz and Boyd, 2005], and can be considered to be the simplest form of Linearly Constrained Minimum Variance (LCMV) beamforming [Frost, 1972; Benesty et al., 2007; Habets et al., 2009] — forcing a single constraint.

For the MVDR case, the objective is to minimise the interference plus noise subject to the single distortionless desired signal constraint. Mathematically this can be expressed as

$$\min \quad \mathbf{w}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \boldsymbol{\psi}_{\mathrm{s}} = 1 \tag{2.52}$$

where $\mathbf{R}_{\mathrm{v}}$ denotes the interference spatial correlation matrix, which can be defined from the definition of the array signals in (2.46) as

$$\mathbf{R}_{\mathrm{v}} = E\left\{ \left(\sum_{n=1}^{N} v_n \boldsymbol{\psi}_{\mathrm{v}_n} \right) \left(\sum_{n=1}^{N} v_n \boldsymbol{\psi}_{\mathrm{v}_n} \right)^H \right\} \tag{2.53}$$

$\mathbf{R}_{\mathrm{n}}$ the sensor noise correlation,

$$\mathbf{R}_{\mathrm{n}} = E\{\mathbf{n}\mathbf{n}^H\} \tag{2.54}$$

and $\boldsymbol{\psi}_{\mathrm{s}}$ the $M \times 1$ transfer function vector describing source to M microphone acoustic transfer functions.

Using the method of Lagrange multipliers this can be formulated as

$$\mathcal{L} = \mathbf{w}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{w} + \lambda \left(1 - \mathbf{w}^H \boldsymbol{\psi}_{\mathrm{s}}\right) \tag{2.55}$$

The beamformer solution can be found by taking the derivative of $\mathcal{L}$ with respect to $\mathbf{w}^H$ and setting this to zero

$$\bigtriangledown_{\mathbf{w}^H} \mathcal{L} = \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{w} - \lambda \boldsymbol{\psi}_{\mathrm{s}} = 0 \tag{2.56}$$

$$\mathbf{w} = \lambda \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\psi}_{\mathrm{s}} \tag{2.57}$$

The value of λ can be obtained by applying the distortionless constraint $\mathbf{w}^H \boldsymbol{\psi}_{\mathrm{s}} = 1$

$$1 = \mathbf{w}^H \boldsymbol{\psi}_{\mathrm{s}} = \lambda \boldsymbol{\psi}_{\mathrm{s}}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\psi}_{\mathrm{s}} \tag{2.58}$$

$$\lambda = \frac{1}{\boldsymbol{\psi}_{\mathrm{s}}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\psi}_{\mathrm{s}}} \tag{2.59}$$

Thus, the MVDR solution is

$$\mathbf{w} = \frac{\left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\psi}_{\mathrm{s}}}{\boldsymbol{\psi}_{\mathrm{s}}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\psi}_{\mathrm{s}}} \tag{2.60}$$

Now suppose multiple transfer function vectors corresponding to either desired signals or interferers were known. As before in the least squares beamformer, the weight solution must satisfy a set of constraints,

$$\boldsymbol{\Psi}^H \mathbf{w} = \mathbf{c} \tag{2.61}$$

where Ψ is an $M \times N$ matrix containing the N known transfer function vectors, and $\mathbf{c}$ is a $N \times 1$ vector containing the constraints to impose.

The LCMV beamforming weight problem can be specified as minimising the interference/noise subject to the constraints given in (2.61)

$$\min \quad \mathbf{w}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{w} \quad \text{s.t.} \quad \boldsymbol{\Psi}^H \mathbf{w} = \mathbf{c} \tag{2.62}$$

where, as before, $\mathbf{R}_{\mathrm{v}}$ is the interference spatial correlation matrix, and $\mathbf{R}_{\mathrm{n}}$ is the sensor noise spatial correlation matrix.

The Lagrangian can be defined as

$$\mathcal{L} = \mathbf{w}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{w} + (\mathbf{c}^H - \mathbf{w}^H \boldsymbol{\Psi}) \boldsymbol{\lambda} \tag{2.63}$$

where $\boldsymbol{\lambda}$ is an $N \times 1$ set of multipliers.

Differentiating $\mathcal{L}$ with respect to $\mathbf{w}^H$ and setting to zero gives

$$\bigtriangledown_{\mathbf{w}^H} \mathcal{L} = \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{w} - \boldsymbol{\Psi} \boldsymbol{\lambda} \tag{2.64}$$

$$\mathbf{w} = \left(\left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\Psi}\right) \boldsymbol{\lambda} \tag{2.65}$$

To satisfy the original constraint condition, $\boldsymbol{\lambda}$ must be chosen such that $\boldsymbol{\Psi}^H \mathbf{w} = \mathbf{c}$:

$$\mathbf{c} = \boldsymbol{\Psi}^H \mathbf{w} = \left(\boldsymbol{\Psi}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\Psi}\right) \boldsymbol{\lambda}$$

$$\boldsymbol{\lambda} = \left(\boldsymbol{\Psi}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\Psi}\right)^{-1} \mathbf{c} \tag{2.66}$$

Therefore the LCMV beamformer solution can be expressed as

$$\mathbf{w} = \left(\left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\Psi}\right) \left(\boldsymbol{\Psi}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\Psi}\right)^{-1} \mathbf{c} \tag{2.67}$$

2.3.4 Generalised Sidelobe Canceller

The generalised sidelobe canceller (GSC) [Griffiths and Jim, 1982; Buckley and Griffiths, 1986] is a technique used to implement LCMV/MVDR beamformers based on the concept of splitting the beamforming task into two beamformers: the first a constraint space beamformer $\mathbf{w}_{\mathrm{c}}$ designed using *known* spatial information (known source and interferer locations); the second an orthogonal space beamformer $\mathbf{w}_{\mathrm{c}\perp}$ designed to remove the *unknown* spatial information (signals from unknown interferer locations).

$$\mathbf{w}_{\mathrm{GSC}} = \mathbf{w}_{\mathrm{c}} - \mathbf{w}_{\mathrm{c}\perp} \tag{2.68}$$

In (2.48) a constraint equation was defined for the application of interference suppression. The objective was to design a beamformer which directed the response to the desired source while simultaneously blocking a set of known interferers. The least squares solution was given as

$$\mathbf{w}_{\mathrm{c}} = \left(\mathbf{\Psi}\mathbf{\Psi}^{H}\right)^{-1}\mathbf{\Psi}\mathbf{c} \tag{2.69}$$

where $\mathbf{\Psi}$ is the $M \times N$ constraint matrix describing the known N transfer function vectors for the known desired source(s) and interferer(s), and $\mathbf{c}$ is the constraint vector containing the N constraints for the corresponding transfer function vector.

The orthogonal beamformer (in the frequency domain) is typically implemented as an adaptive two-stage system (depicted in Figure 2.5)

$$\mathbf{w}_{c\perp} = \mathbf{B}\mathbf{w}_{\mathrm{ad.}} \tag{2.70}$$

where $\mathbf{B}$ is the $M \times (M-N)$ blocking matrix (whose columns are orthogonal to those in $\mathbf{\Psi}$), and $\mathbf{w}_{\mathrm{ad.}}$ is an $(M-N) \times 1$ adaptive filter vector, which

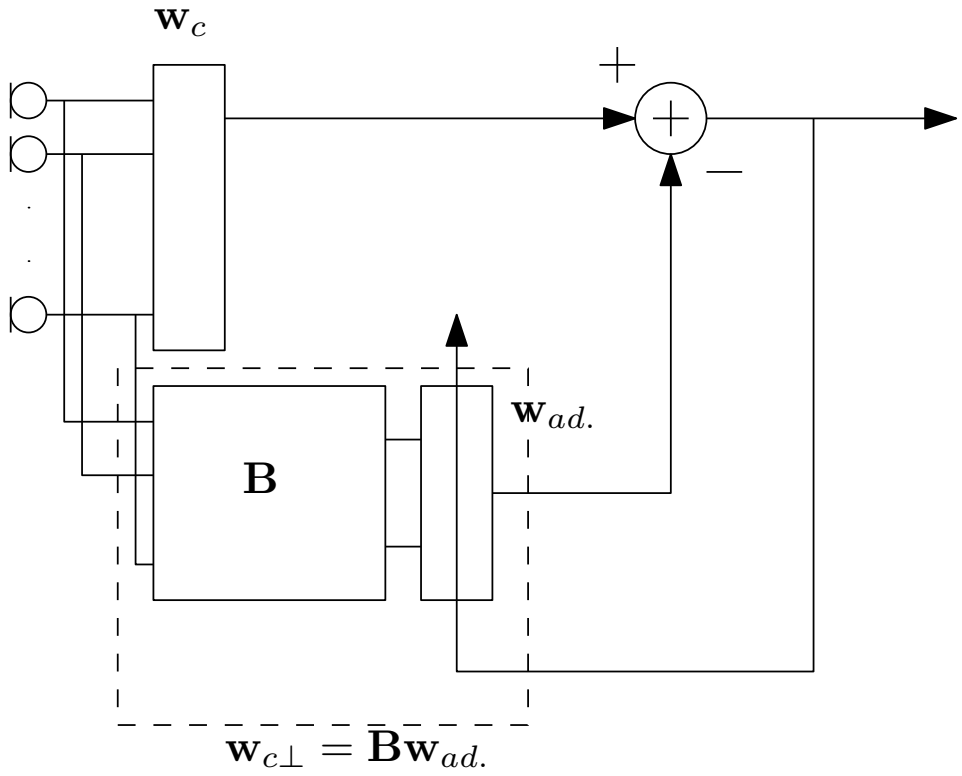

Figure 2.5: A simple example of a generalised sidelobe canceller.

is updated over time for each frequency index (in the short time-frequency domain) as [Griffiths and Jim, 1982; Van Trees, 2004]

$$\mathbf{w}_{\text{ad.}}^{+} = \mathbf{w}_{\text{ad.}} + \mu \left(\mathbf{w}_{\text{c}}^{H} \mathbf{x} - \mathbf{w}_{\text{ad.}}^{H} (\mathbf{B}\mathbf{x}) \right) \mathbf{x} \tag{2.71}$$

where μ is the adaptive filter step-size. The blocking matrix prevents the signals aligned with the constraint matrix from travelling down the lower path of Figure 2.5, allowing (in free-field conditions, ignoring reverberation) just unknown signals to pass through. The unknown signals may also be partly aligned with the constraint matrix contributing to unwanted interference/noise

in the upper path output. The adaptive filter in the lower path is able to compensate for this by removing the correlated components between the two paths.

A technique (but not the only technique) for computing the orthonormal blocking matrix is to perform singular value decomposition (or eigenvalue decomposition) on the $M \times M$ Hermitian spatial correlation matrix

$$\begin{aligned} \mathbf{R}_{\mathrm{c}} &= \boldsymbol{\psi}\boldsymbol{\psi}^H \\ &= \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^H \end{aligned} \tag{2.72}$$

where $\mathbf{U}$ is an orthonormal $N \times N$ matrix of left-singular vectors, $\boldsymbol{\Sigma}$ is a diagonal matrix containing the singular values, and $\mathbf{V}$ is an orthonormal matrix of right-singular vectors. As $\mathbf{R}_{\mathrm{c}}$ is Hermitian, the $\mathbf{U}$ and $\mathbf{V}$ matrices are identical.

The blocking matrix can be formed as column matrix populated using the singular vectors in $\mathbf{U}$ corresponding to the zero-valued diagonal entries of $\boldsymbol{\Sigma}$. It should be noted that if $\mathbf{R}_{\mathrm{c}}$ is full-rank, no orthogonal blocking matrix exists — and the least squares solution in (2.50) is already optimal.

In practice, the blocking matrix formulation is difficult to design as it relies on *perfect* constraint matrix knowledge, which is an unrealistic assumption in most scenarios. If the blocking matrix does not perfectly null a desired signal, it will leak into the blocking path resulting in desired signal attenuation, as the adaptive filter is designed to remove the correlated components between the upper and lower parts in Figure 2.5. In order to accurately obtain the blocking matrix, additional processing steps are required: for example identifying the directional of arrival of the desired source (or interferers), or voice/signal activity detection methods [Ramirez et al., 2004] used to obtain the blocking

matrix during signal silence periods.

2.3.5 Multichannel Wiener Filtering

The MVDR and LCMV beamformer methods optimise the weights usually to satisfy a distortionless response constraint to the desired source. Further optimisation of SINR can be achieved by relaxing the distortionless response criteria and considering a complete least squares solution — the previous MVDR/LCMV solutions in (2.60, 2.67) only considered the transfer function information; here we will consider the actual desired/interferer signal information along with the spatial information.

Defining the error function (at a single frequency) as

$$\epsilon = \mathbf{w}^H \mathbf{x} - d \tag{2.73}$$

where $\mathbf{w}$ is the beamformer weight vector to be obtained, $\mathbf{x}$ is the vector of signals received at the array, and d denotes the desired (reference) signal. The reference signal is usually taken as the desired signal received at one of the microphones in the array.

A least squares weight solution for this problem can be found by minimising the cost function

$$J = E\{|\epsilon|^2\} = (\mathbf{w}^H \mathbf{x} - d)(\mathbf{x}^H \mathbf{w} - d^*) \tag{2.74}$$

$$= \mathbf{w}^H E\{\mathbf{x}\mathbf{x}^H\}\mathbf{w} - 2E\{d\mathbf{w}^H \mathbf{x}\} + E\{dd^*\} \tag{2.75}$$

The weight vector which minimises J can be obtained by differentiating J with respect to $\mathbf{w}^H$ and finding the minimum gradient solution

$$\bigtriangledown_{\mathbf{w}^H} J = 2\mathbf{R}_x \mathbf{w} - 2\mathbf{r}_{d\mathbf{x}} \tag{2.76}$$

where

$$\mathbf{R}_x = E\{\mathbf{x}\mathbf{x}^H\} \tag{2.77}$$

denotes the array input correlation matrix and

$$\mathbf{r}_{d\mathbf{x}} = E\{d\mathbf{x}\} \tag{2.78}$$

denotes the array input to the desired signal cross-correlation vector. Assuming the desired signal and interference are uncorrelated, the cross-correlation can be expressed as

$$\mathbf{r}_{d\mathbf{x}} = \sigma_s^2 \boldsymbol{\psi}_{\mathrm{s}} \tag{2.79}$$

The multichannel Wiener filter [Van Trees, 2004] solution is therefore

$$\mathbf{w}_{\mathrm{MWF}} = \mathbf{R}_x^{-1}\mathbf{r}_{d\mathbf{x}} \tag{2.80}$$

This can be shown be decomposable into two parts: the MVDR beamformer plus a single channel Wiener filter post-processor.

Assuming the desired source transfer function vector is deterministic, the array input correlation matrix can be expanded as

$$\mathbf{R}_x = \sigma_s^2 \boldsymbol{\psi}_{\mathrm{s}} \boldsymbol{\psi}_{\mathrm{s}}^H + \mathbf{R}_{\mathrm{v}} \tag{2.81}$$

Noting that the desired signal component is a rank-1 update, the Sherman-Morrison formula [Hager, 1989; Van Trees, 2004] can be used to derive a more useful form of the inverse

$$\mathbf{R}_x^{-1} = \mathbf{R}_{\mathrm{v}}^{-1} - \frac{\mathbf{R}_{\mathrm{v}}^{-1}\sigma_s^2 \boldsymbol{\psi}_{\mathrm{s}} \boldsymbol{\psi}_{\mathrm{s}}^H \mathbf{R}_{\mathrm{v}}^{-1}}{1 + \sigma_s^2 \boldsymbol{\psi}_{\mathrm{s}}^H \mathbf{R}_{\mathrm{v}}^{-1} \boldsymbol{\psi}_{\mathrm{s}}} \tag{2.82}$$

Defining

$$\lambda^2 = \boldsymbol{\psi}_{\mathrm{s}}^H \mathbf{R}_{\mathrm{v}}^{-1} \boldsymbol{\psi}_{\mathrm{s}} \tag{2.83}$$

(2.82) becomes

$$\mathbf{R}_x^{-1} = \mathbf{R}_{\mathrm{v}}^{-1} - \frac{\mathbf{R}_{\mathrm{v}}^{-1}\sigma_s^2\boldsymbol{\psi}_{\mathrm{s}}\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}}{1+\lambda^2\sigma_s^2} \tag{2.84}$$

Inserting (2.84) into (2.80) gives

$$\mathbf{w}_{\mathrm{MWF}} = \sigma_s^2\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}\left(1 - \frac{\lambda^2\sigma_s^2}{1+\lambda^2\sigma_s^2}\right) \tag{2.85}$$

The bracketed term can be simplified as follows

$$\mathbf{w}_{\mathrm{MWF}} = \sigma_s^2\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}\left(\frac{1+\lambda^2\sigma_s^2}{1+\lambda^2\sigma_s^2} - \frac{\lambda^2\sigma_s^2}{1+\lambda^2\sigma_s^2}\right) \tag{2.86}$$

$$= \sigma_s^2\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}\left(\frac{1}{1+\lambda^2\sigma_s^2}\right) \tag{2.87}$$

$$= \sigma_s^2\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}\left(\frac{1}{\lambda^2}\frac{1}{(\lambda^{-2}+\sigma_s^2)}\right) \tag{2.88}$$

$$= \lambda^{-2}\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}\left(\frac{\sigma_s^2}{\lambda^{-2}+\sigma_s^2}\right) \tag{2.89}$$

$$= \frac{\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}}{\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}}\left(\frac{\sigma_s^2}{\sigma_s^2+\lambda^{-2}}\right) \tag{2.90}$$

The vector term is the MVDR beamformer as previously derived in an earlier section, and the scalar term is equivalent to the single channel Wiener filter [Jeub and Vary, 2010], where the λ^{-2} term denotes the output interferer power after MVDR beamforming. This can be demonstrated by considering only the MVDR portion applied to the array outputs

$$\mathbf{w}_{\mathrm{MVDR}}^H\mathbf{x} = d\frac{\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}}{\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}} + \frac{\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\mathbf{v}}{\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}} \tag{2.91}$$

Squaring and taking the expectation

$$E\{\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}\} = \sigma_{\mathrm{s}}^2 + \frac{\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}E\{\mathbf{v}\mathbf{v}^H\}\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}}{\left(\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}\right)^2} \tag{2.92}$$

Noting that

$$\mathbf{R}_{\mathrm{v}} = E\{\mathbf{v}\mathbf{v}^H\} \tag{2.93}$$

the output power after MVDR beamforming is

$$E\{\mathbf{w}^H\mathbf{x}\mathbf{x}^H\mathbf{w}\} = \sigma_{\mathrm{s}}^2 + \frac{\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\mathbf{R}_{\mathrm{v}}\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}}{\left(\boldsymbol{\psi}_{\mathrm{s}}^H\mathbf{R}_{\mathrm{v}}^{-1}\boldsymbol{\psi}_{\mathrm{s}}\right)^2} = \sigma_{\mathrm{s}}^2 + \frac{\lambda^2}{\lambda^4} = \sigma_{\mathrm{s}}^2 + \lambda^{-2} \qquad (2.94)$$

where σ_s^2 denotes the expected desired signal variance, and λ^{-2} denotes the expected output interference variance.

It can be seen after further factorisation that the single channel Wiener filter (SWF) can be computed if the interference/noise to signal ratio (or one over the signal to interferer/noise (SINR)) is known.

$$\mathrm{SWF} = \frac{\sigma_s^2}{\sigma_s^2 + \lambda^{-2}} = \frac{\sigma_s^2}{\sigma_s^2}\frac{1}{1 + \frac{\lambda^{-2}}{\sigma_s^2}} = \frac{1}{1 + \mathrm{SINR}^{-1}} \qquad (2.95)$$

The multichannel Wiener filter can be computed provided the location of the desired signal (providing the transfer function vector $\boldsymbol{\psi}_{\mathrm{s}}$), the interference correlation matrix $\mathbf{R}_{\mathrm{v}}$, and MVDR beamformed output SIR/SINR are known.

2.3.6 Maximum SINR Beamforming

An alternative optimal beamformer can be derived by considering the weights required to maximise the signal to interferer plus noise ratio (SINR). The output of a beamformed array can be expressed as

$$y(i,k) = s(i,k)\mathbf{w}^H(i,k)\boldsymbol{\psi}_{\mathrm{s}}(k) + \mathbf{w}^H(i,k)\mathbf{v}(i,k) + \mathbf{w}^H(i,k)\mathbf{n}(i,k) \qquad (2.96)$$

Omitting time/wavenumber indexing for clarity, the expected output power of the beamformed array can be expressed as

$$E\{yy^*\} = \sigma_{\mathrm{s}}^2\mathbf{w}^H E\{\boldsymbol{\psi}_{\mathrm{s}}\boldsymbol{\psi}_{\mathrm{s}}^H\}\mathbf{w} + \mathbf{w}^H E\{\mathbf{v}\mathbf{v}^H\}\mathbf{w} + \mathbf{w}^H E\{\mathbf{n}\mathbf{n}^H\}\mathbf{w} \qquad (2.97)$$

assuming the desired signal, interferers and noise are uncorrelated.

The maximum SINR beamformer [Shahbazpanahi et al., 2003] can be described as a set of beamforming weights $\mathbf{w}$ which maximises

$$\text{SINR} = \frac{\mathbf{w}^H E\{\boldsymbol{\psi}_\text{s}\boldsymbol{\psi}_\text{s}^H\}\mathbf{w}}{\mathbf{w}^H E\{\mathbf{v}\mathbf{v}^H\}\mathbf{w} + \mathbf{w}^H E\{\mathbf{n}\mathbf{n}^H\}\mathbf{w}} \tag{2.98}$$

where we have ignored the desired signal variance as in general this is not known in advance.

The weights can be solved for by minimising the output interference plus noise power such that the desired signal power is equal to some arbitrary constraint:

$$\min \quad \mathbf{w}^H \left[\mathbf{R}_\text{v} + \mathbf{R}_\text{n}\right]\mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H\mathbf{R}_\text{s}\mathbf{w} = \alpha \tag{2.99}$$

where $\mathbf{R}_\text{s} = E\{\boldsymbol{\psi}_\text{s}\boldsymbol{\psi}_\text{s}^H\}$ denotes the desired source spatial correlation matrix, $\mathbf{R}_\text{v} = E\{\mathbf{v}\mathbf{v}^H\}$ denotes the interference spatial correlation matrix, and $\mathbf{R}_\text{n} = E\{\mathbf{n}\mathbf{n}^H\}$ denotes the sensor noise spatial correlation matrix.

Equation (2.99) can be solved using the Lagrange multiplier method

$$\mathcal{L} = \mathbf{w}^H \left[\mathbf{R}_\text{v} + \mathbf{R}_\text{n}\right]\mathbf{w} + \lambda\left(\alpha - \mathbf{w}^H\mathbf{R}_\text{s}\mathbf{w}\right) \tag{2.100}$$

Differentiating $\mathcal{L}$ with respect to $\mathbf{w}^H$,

$$\bigtriangledown_{\mathbf{w}^H}\mathcal{L} = \left[\mathbf{R}_\text{v} + \mathbf{R}_\text{n}\right]\mathbf{w} - \lambda\mathbf{R}_\text{s}\mathbf{w} \tag{2.101}$$

A solution for $\mathbf{w}$ can be obtained by setting the gradient $\bigtriangledown_{\mathbf{w}^H}\mathcal{L}$ to zero. Re-arranging (2.101) leads to the generalised eigenvalue equation

$$\left[\mathbf{R}_\text{v} + \mathbf{R}_\text{n}\right]\mathbf{w} = \lambda\mathbf{R}_\text{s}\mathbf{w} \tag{2.102}$$

where the beamformer solution $\mathbf{w}$ is attained by choosing the eigenvector associated with the *largest* eigenvalue.

Discussion

The maximum SINR beamformer formulation, as it will be seen in later chapters, can be a powerful method for designing spatially robust beamformers if the desired signal spatial correlation matrix $\mathbf{R}_{\text{s}}$ is designed appropriately. The formulation also easily leads to the design of robust blocking beamformers (nullformers) — either by switching $\mathbf{R}_{\text{s}}$ for $\mathbf{R}_{\text{v}}$ in (2.102), or selecting the eigenvector associated with the *smallest* eigenvalue.

2.3.7 Spatial Correlation Modelling

Most of the beamforming algorithms in previous subsections rely on knowledge of the interference correlation matrix $\mathbf{R}_{\text{v}}$. A simple and common model to use is to assume that the interference is fixed, i.e., is not time/space varying. In this model it is possible to pre-compute estimated correlation matrices relatively easily.

Isotropic Interference

A simple assumption for fixed beamformer design is to assume that interference arrives at the array evenly from all possible directions. This assumption is commonly used to model high levels of reverberation [Schwarz and Kellermann, 2015], where large numbers of reflections can be approximated as an isotropic distribution of sources.

A spatial correlation function due to a source at some distant position (r, θ, ϕ) incident on two sensors located at (r_a, θ_a, ϕ_a) and (r_b, θ_b, ϕ_b), close to the coordinate origin, can be defined as

$$\mathbf{R}_{\text{v}}[a, b] = \psi_a(r, r_a, \Omega, \Omega_a)\psi_b^*(r, r_b, \Omega, \Omega_b) \tag{2.103}$$

where Ω denotes shorthand for the angles (θ, ϕ) the ψ functions are defined using the near-field point source description in (2.15) as

$$\psi_a(r, r_a, \Omega, \Omega_a) = \frac{e^{ik|r_a - r|}}{4\pi|r_a - r|} \tag{2.104}$$

The ψ functions can be described in terms of spherical Bessel functions and spherical harmonics using the spherical Bessel function addition properties [Abramowitz and Stegun, 1964, (10.1.45, 10.1.46)]

$$\frac{e^{ikR}}{4\pi R} = ik \sum_{n=0}^{\infty} (2n+1) j_n(kr_a) h_n(kr) P_n(\cos\Omega_0) \tag{2.105}$$

where

$$R = \sqrt{r^2 + r_a^2 - 2rr_a \cos\Omega_0} \tag{2.106}$$

$$\cos\Omega_0 = \cos\theta\cos\theta_a + \sin\theta\sin\theta_a\cos(\phi - \phi_a) \tag{2.107}$$

The spherical harmonic addition theorem [Clapp, 1970]

$$\frac{(2n+1)}{4\pi} P_n(\cos\Omega_0) = \sum_{m=-n}^{n} Y_n^m(\Omega) Y_n^{m*}(\Omega_a) \tag{2.108}$$

can be used to expand (2.104) into

$$\psi_a = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{n} j_n(kr_a) h_n(kr) Y_n^m(\Omega) Y_n^{m*}(\Omega_a) \tag{2.109}$$

Extending the correlation function to an infinite number of sources over a spherical volume gives the integral expression

$$\mathbf{R}_{\mathrm{v}}[a, b] = \int_{\mathrm{vol.}} \psi_a(r, r_a, \Omega, \Omega_a) \psi_b^*(r, r_b, \Omega, \Omega_b)\, dV \tag{2.110}$$

A simple model of isotropic interference is to assume that the interferers are in the far-field, i.e., at some fixed distance r much greater than the microphone

radii r_a and r_b. In this case, the volume integral can be reduced to a surface integral. Substituting (2.109) for the ψ functions in (2.110) leads to the expression

$$\mathbf{R}_{\mathrm{v}}[a,b] = k^2 \sum_{n_a,n_b} \sum_{m_a,m_b} j_{n_a}(kr_a) j_{n_b}(kr_b) h_{n_a}(kr) h^*_{n_b}(kr) \\ Y^{m_a *}_{n_a}(\Omega_a) Y^{m_b}_{n_b}(\Omega_b) \int_\Omega Y^{m_a}_{n_a}(\Omega) Y^{m_b *}_{n_b}(\Omega)\, d\Omega \tag{2.111}$$

Using the orthonormality property of the spherical harmonics

$$\int_\Omega Y^{m_1}_{n_1}(\Omega) Y^{m_2 *}_{n_2}(\Omega)\, d\Omega = \delta_{n_1,n_2,m_1,m_2} \tag{2.112}$$

the quad summation in (2.111) reduces to the simpler form

$$\mathbf{R}_{\mathrm{v}}[a,b] = k^2 \sum_n \sum_m j_n(kr_a) j_n(kr_b) h_n(kr) h^*_n(kr) Y^{m*}_n(\Omega_a) Y^m_n(\Omega_b) \tag{2.113}$$

where n and m are the common order and degree terms arising from the orthogonality relation in (2.112). If the interference is assumed to originate from a distance r much greater than the microphone radii, the spherical Hankel function approximation [Williams, 1999, (6.68)]

$$h_n(kr) \simeq i^n \frac{e^{ikr}}{kr} \quad \text{for} \quad kr \gg 1 \tag{2.114}$$

can be substituted into (2.113) to obtain

$$\mathbf{R}_{\mathrm{v}}[a,b] = \frac{1}{r^2} \sum_n \sum_m j_n(kr_a) j_n(kr_b) Y^{m*}_n(\Omega_a) Y^m_n(\Omega_b) \tag{2.115}$$

The spherical harmonic addition theorem in (2.108) can be used to simplify further to

$$\mathbf{R}_{\mathrm{v}}[a,b] = \frac{1}{4\pi r^2} \sum_n (2n+1) j_n(kr_a) j_n(kr_b) P_n(\cos \Omega_{a,b}) \tag{2.116}$$

where

$$\cos\Omega_{a,b} \triangleq \cos\theta_a \cos\theta_b + \sin\theta_a \sin\theta_b \cos(\phi_a - \phi_b) \tag{2.117}$$

Using the spherical Bessel function addition theorem [Abramowitz and Stegun, 1964, (10.1.45)], the correlation function can be expressed as

$$\mathbf{R}_{\mathrm{v}}[a,b] = \frac{j_0(kr_{a,b})}{4\pi r^2} \tag{2.118}$$

where $4\pi r^2$ is a simple interferer distance normalisation factor, and $r_{a,b}$ denotes the distance between the elements of the microphone pair. In the literature, the isotropic correlation function is usually given without the normalisation factor as [Piersol, 1978; Schwarz and Kellermann, 2015]

$$\mathbf{R}_{\mathrm{v}}[a,b] = j_0(kr_{a,b}) = \frac{\sin(kr_{a,b})}{kr_{a,b}} \tag{2.119}$$

Anisotropic Interference

An anisotropic interferer distribution can be modelled by modifying (2.110) to include some kind of selection function or probability density function,

$$\mathbf{R}_{\mathrm{v}} = \int_{\mathrm{vol.}} p(r,\Omega)\psi_a(r,r_a,\Omega,\Omega_a)\psi_b^*(r,r_a,\Omega,\Omega_b)\,dV \tag{2.120}$$

In Chapters 4 and 5, various correlation functions are derived for anisotropic interference and/or desired source distribution modelling.

2.3.8 Beamforming Issues

All of the interference cancelling methods outlined in the previous subsections rely on some knowledge of the transfer functions (or their statistics) from the interferers to the array and/or precise knowledge of the desired source to array transfer functions. In many scenarios the desired source to array

transfer functions can be estimated reasonably easily as the position of the desired source may be well known, for example, a microphone array in a cellphone could make a reasonably safe assumption that the user's mouth would normally be on the front-side and near the bottom of the phone; similarly, a microphone array built into a laptop computer could be designed to respond to input directly in front of the screen. Additionally, methods which estimate the relative transfer function [Cohen, 2004; Talmon et al., 2009] (just the propagation between the microphones) of the desired source can and have been used to adaptively optimise the beamformer [Gannot et al., 2001; Gannot and Cohen, 2002] under certain conditions.

In good SIR/SINR environments, voice activity detection can be used to detect gaps in speech which provides an opportunity to compute the interferer correlation matrix and/or SINR. This method is commonly used to implement the adaptive versions of MVDR [Ba et al., 2007; Chen and Benesty, 2011; Cauchi et al., 2014], LCMV [Chen and Benesty, 2013], GSC, multichannel Wiener filters [Van den Bogaert et al., 2009] and similar algorithms [Doclo and Moonen, 2002]. The drawback of this technique is the requirement of a good voice activity detector, i.e., one with good voice/signal detection (few false positives/negatives) and robustness to high noise environments. Additionally, voice activity detection has the disadvantage of assuming there is only one talker, which may be an unrealistic assumption for many speech applications, where there may be one or more 'interfering' talkers near the microphone array.

2.3.9 Beamformer Robustness

Spatial Robustness

For the intended application of this thesis, the beamformer designs must be spatially robust. That is, the designs should assume imperfect knowledge of the target position relative to the array, and be able to tolerate this scenario. Similarly to the interference modelling in Section 2.3.7 (2.120), a spatially robust beamformer can be designed by assuming a distribution of possible target directions/locations, and designing spatial correlation functions as appropriate. The maximum SINR beamformer in Section 2.3.6 provides a method for designing the spatially robust beamformers once the correlation functions have been computed.

Numerical Robustness

An important issue in beamformer design is numerical robustness — the ability to tolerate random errors in the system, such as microphone calibration errors, and sensor noise.

The white noise gain [Cox et al., 1986] of an array is a measure of its robustness to intrinsic microphone errors (array position error, frequency response, etc.), modelled as white noise:

$$\text{WNG} = \frac{\mathbf{w}^H \mathbf{R}_{\text{s}} \mathbf{w}}{\mathbf{w}^H \mathbf{w}} \tag{2.121}$$

Low white noise gain is indicative of sensitivity to errors, and conversely high white noise gain indicates the ability to tolerate these errors, which is a useful property for practical arrays.

Improvements to numerical robustness of the beamformer solutions can be

achieved by increasing the sensor noise variance used to design the diagonal sensor noise spatial correlation matrix.

2.4 Blind Source Separation

Blind source separation algorithms are a class of signal processing algorithms which attempt to blindly identify mixtures of signals, where the mixing process between the source(s) and the microphones is unknown or limited knowledge is available. The mechanism for identifying the separation filters is often based on the concept of minimising mutual information between the output signals.

The classic problem which introduces BSS is the cocktail party problem in which there are multiple simultaneous talkers and listeners [Haykin and Chen, 2005] and the objective is to extract a desired speech signal from the mixture.

Starting with a basic instantaneous model, the problem can be stated as

$$\begin{bmatrix} x_1(t) \\ x_2(t) \\ \vdots \\ x_M(t) \end{bmatrix} = \begin{bmatrix} a_{1,1} & a_{1,2} & \dots & a_{1,M} \\ a_{2,1} & a_{2,2} & \dots & a_{2,M} \\ \vdots & & \ddots & \vdots \\ a_{M,1} & a_{M,2} & \dots & a_{M,N} \end{bmatrix} \begin{bmatrix} s_1(t) \\ s_2(t) \\ \vdots \\ s_N(t) \end{bmatrix} + \begin{bmatrix} n_1(t) \\ n_2(t) \\ \vdots \\ n_N(t) \end{bmatrix} \tag{2.122}$$

$$\mathbf{x}(t) = \mathbf{A}\mathbf{s}(t) + \mathbf{n}(t) \tag{2.123}$$

where $\mathbf{x}$ is the set of sensor observations, $\mathbf{A}$ a matrix describing the signal to sensor mixture model, $\mathbf{s}$ the set of signals (speech for example), and $\mathbf{n}$ the sensor noise.

The objective can be stated as finding some demixing matrix (transform) $\mathbf{B}$ which separates the mixed signals. This demixing matrix can be obtained

through various techniques such as Principal Component Analysis (PCA) [Jolliffe, 2002] and Independent Component Analysis (ICA) [Hyvärinen et al., 2004]. In PCA, the objective is to find some transform which decorrelates the outputs, under the assumption that the original sources are not correlated (and have zero mean). In ICA the objective is to find some transform which makes the outputs statistically independent, usually by assuming non-Gaussianity, non-whiteness, and/or non-stationarity of the sources.

The PCA transform can be computed by constructing an input data covariance matrix

$$\mathbf{R}_{xx} = E\{\mathbf{x}\mathbf{x}^H\} \tag{2.124}$$

$$= \mathbf{Q}\mathbf{\Lambda}\mathbf{Q}^H \quad (EVD) \tag{2.125}$$

$$= \mathbf{U}\mathbf{\Sigma}^2\mathbf{U}^H \quad (SVD) \tag{2.126}$$

and computing either the eigenvalue decomposition or singular value decomposition to extract the eigenvector/singular-vector matrix which when applied to $\mathbf{x}$, separates the original signals.

$$\tilde{\mathbf{s}} = \mathbf{Q}^H\mathbf{x} \quad (EVD) \tag{2.127}$$

$$= \mathbf{U}^H\mathbf{x} \quad (SVD) \tag{2.128}$$

The ICA transforms such as JADE (Joint Approximation Diagonalization of Eigen-matrices) [Cardoso and Souloumiac, 1993] and FastICA [Hyvärinen and Oja, 1997; Hyvärinen, 2001] typically utilise PCA, to decorrelate the inputs, followed by processing using higher order statistics (such as kurtosis) to design filters which minimise the Gaussianity of the outputs.

In FastICA for example, the ICA transform is computed by pre-whitening the input data using PCA

$$\tilde{\mathbf{x}} = \mathbf{Q}^H\mathbf{x} \tag{2.129}$$

and finding a set of demixing filters — column vectors of a matrix $\mathbf{W}$ which when applied to the whitened input, separates out the original signals.

The demixing filters can be found by minimising/maximising some cost function which approximates, for example, the kurtosis of the demixed signal.

$$J = E\{(\mathbf{w}_i^H \tilde{\mathbf{x}})^4\} - 3E\{(\mathbf{w}_i^H \tilde{\mathbf{x}})^2\}^2 \tag{2.130}$$

A gradient descent/ascent method can be used to find the minimum/maximum kurtosis value. Taking the gradient of (2.130) with respect to $\mathbf{w}_i^H$ gives

$$J = E\{\mathbf{w}_i^H \tilde{\mathbf{x}}\tilde{\mathbf{x}}^H \mathbf{w}_i \mathbf{w}_i^H \tilde{\mathbf{x}}\tilde{\mathbf{x}}^H \mathbf{w}_i\} - 3E\{\mathbf{w}_i^H \tilde{\mathbf{x}}\tilde{\mathbf{x}}^H \mathbf{w}_i\}^2 \tag{2.131}$$

$$\bigtriangledown_{\mathbf{w}_i^H} J = E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H \mathbf{w}_i \mathbf{w}_i^H \tilde{\mathbf{x}}\tilde{\mathbf{x}}^H \mathbf{w}_i\} - 3\mathbf{w}_i \mathbf{w}_i^H \mathbf{w}_i \tag{2.132}$$

$$\bigtriangledown_{\mathbf{w}_i^H} J = E\{\tilde{\mathbf{x}}(\mathbf{w}_i^H \tilde{\mathbf{x}})^3\} - 3\mathbf{w}_i \tag{2.133}$$

where it has been assumed that the demixing filters are of unit norm ($\mathbf{w}_i^H \mathbf{w}_i = 1$), and the whitening process diagonalises $E\{\tilde{\mathbf{x}}\tilde{\mathbf{x}}^H\}$.

The weights for each demixing filter, corresponding to each output signal, can be computed iteratively as

$$\mathbf{w}_{i+} = \mathbf{w}_i - \mu \left(E\{\tilde{\mathbf{x}}(\mathbf{w}_i^H \tilde{\mathbf{x}})^3\} - 3\mathbf{w}_i \right) \tag{2.134}$$

where μ denotes the gradient ascent/descent parameter.

During convergence, the filter is orthogonalised with respect to the other filters in the demixing matrix $\mathbf{W}$, using the Gram-Schmidt process for example.

The iterative steps continue until the filter has converged, which can be determined by computing the inner product of the current and previous filter vectors. Convergence occurs when the inner product is 1 (or within some tolerance ϵ).

Once the demixing matrix has been computed, the separated signals can be obtained by calculating

$$\mathbf{y} = \mathbf{W}^H \mathbf{x} \tag{2.135}$$

The outputs after applying the PCA/ICA transform exhibit channel ordering and scaling ambiguities related to the fact that the mixing matrix and order of the original signals are unknown.

The channel-order and scaling ambiguities lead to issues when attempting to use ICA on convolutive mixtures. Convolutive mixtures require performing multiple instantaneous demixing units, one for each frequency bin of a chosen transform size corresponding to the length of the convolution, which introduces these permutation and scaling errors for each frequency band. Due to the permutation ambiguity in particular, using a frequency bin-wise ICA algorithm is unusable for wideband signals.

Attempts have been made to correct these permutation problems [Sawada et al., 2004] by looking at correlations between the demixing matrices for neighbouring frequency bins, however these methods require introducing an additional analysis step after the demixing matrices have been constructed, increasing the computational cost. Additionally, the method specified in [Sawada et al., 2004] requires the use of a direction of arrival process which may not work in high noise environments where the desired signal exhibits less power than the interferers, a focus of this thesis.

2.4.1 TRINICON

TRINICON (Triple-N independent component analysis for convolutive mixtures) [Buchner et al., 2004a; Buchner et al., 2004b; Kellermann et al., 2006] is

a framework of separation algorithms based on three signal properties — non-gaussianity, non-whiteness and non-stationarity. The objective is to find a set of demixing filters which minimises the Kullback-Leiber distance (equivalent to minimising the mutual information) between the original signals and the output of the demixing system [Buchner et al., 2004b]. This measure requires knowledge of the source probability density functions, which in general are not known in advance. One simplification which can be used [Aichner et al., 2005] is to assume Gaussian signals, simplifying the filter update equations at the cost of demixing performance.

Considering the output auto/cross-correlation matrix

$$\mathbf{R}_{yy} = \mathbf{W}^H \mathbf{R}_{xx} \mathbf{W} = \begin{bmatrix} E\{\mathbf{Y}_1\mathbf{Y}_1^H\} & E\{\mathbf{Y}_1\mathbf{Y}_2^H\} & ... & E\{\mathbf{Y}_1\mathbf{Y}_N^H\} \\ \vdots & \ddots & ... & \vdots \\ E\{\mathbf{Y}_N\mathbf{Y}_1^H\} & E\{\mathbf{Y}_N\mathbf{Y}_2^H\} & ... & E\{\mathbf{Y}_N\mathbf{Y}_N^H\} \end{bmatrix} \quad (2.136)$$

where $\mathbf{Y}_i$ is a Toeplitz matrix of time-domain samples for the i^{th} output channel; the input auto/cross-correlation matrix $\mathbf{R}_{xx}$ is defined as

$$\mathbf{R}_{xx} = \begin{bmatrix} E\{\mathbf{X}_1\mathbf{X}_1^H\} & E\{\mathbf{X}_1\mathbf{X}_2^H\} & ... & E\{\mathbf{X}_1\mathbf{X}_N^H\} \\ \vdots & \ddots & ... & \vdots \\ E\{\mathbf{X}_N\mathbf{X}_1^H\} & E\{\mathbf{X}_N\mathbf{X}_2^H\} & ... & E\{\mathbf{X}_N\mathbf{X}_N^H\} \end{bmatrix} \quad (2.137)$$

(where $\mathbf{X}_i$ is a Toeplitz matrix of time-domain samples for the i^{th} channel); and the BSS filter matrix is defined as a matrix of input-output Sylvester structure convolution demixing filter submatrices

$$\mathbf{W} = \begin{bmatrix} \mathbf{W}_{1,1} & \mathbf{W}_{1,2} & ... & \mathbf{W}_{1,N} \\ \vdots & \ddots & & \vdots \\ \mathbf{W}_{N,1} & \mathbf{W}_{N,2} & ... & \mathbf{W}_{N,N} \end{bmatrix} \quad (2.138)$$

The demixing filters are computed through minimising an appropriate output cross-correlation reduction cost function using a gradient descent method

$$\mathbf{W} = \mathbf{W} - \mu \triangle \mathbf{W} \tag{2.139}$$

where μ is the gradient descent/ascent parameter and $\triangle \mathbf{W}$ is the filter update equation.

The update equation at a time-block index m corresponding to the cost function for the second order statistics case is given in [Buchner et al., 2004a; Buchner et al., 2004b] as

$$\triangle \mathbf{W} = 2 \sum_{i=0}^{\infty} \beta(i,m) \mathbf{W} [\text{offDiag}(\mathbf{R}_{\text{yy}}(i)) \ \text{blockDiag}^{-1}(\mathbf{R}_{\text{yy}}(i))] \tag{2.140}$$

where β represents a window function describing the type of update in operation, and i denotes a previous time-block index. In [Buchner et al., 2004b] three types of update operations are described as offline updates, where the entire signal is processed at once; online, where the filters are continually updated for new samples arriving; and block-online, which combines the two by processing the signal block-wise (as for the update equation given in (2.140)). The blockDiag and offDiag operators are defined as the block-diagonal and off-diagonal matrix operators which select the respective submatrices.

The overall technique described in [Buchner et al., 2004b] from the outset looks to be a computationally intractable problem. For any practical system, the filter lengths required to demix signals would lead to matrix equations involving multiplications with millions of entries to be computed. However, the structure of the matrices described earlier represent convolutions, which can be efficiently computed in the frequency domain via the fast Fourier transform [Aichner et al., 2005].

More recent work using this algorithm has involved efforts to fix the channel ordering permutation problem inherent in blind source algorithms by exploiting partial knowledge of the system to demix. If the source position or direction is known, constraints [Zheng et al., 2009] can be introduced to restrict the filters to essentially beamform towards (or away from) a known source. In this thesis, a similar method in which the beamforming stage occurs before the BSS stage (as opposed to during) is proposed in Chapter 6.

2.5 Conclusions

The various beamforming methods detailed in Sections 2.3.1 to 2.3.6 provide a set of algorithms to enhance signals from a known direction/position. The majority of the existing beamformer techniques rely on (ideally the precise) knowledge of the desired source direction/position to perform well. The maximum SINR beamformer solution presented in Section 2.3.6 provides an optimal general solution provided the desired signal and interferer spatial correlation matrices can be accurately modelled.

Blind source methods provide interesting solutions to speech enhancement by providing mechanisms for blindly identifying signals in mixtures. The disadvantages of these class of techniques is that they require post-processor identification techniques to correct permutation ambiguity issues which can be difficult to implement. Beamformer-type correction methods are relatively simple to implement and work well in solving channel ordering permutations.

The proposed method in Chapter 6 combines spatially robust beamforming/nullforming and 2-channel blind source separation to produce an adaptive interference/noise reduction system. The (assumed to be) non-optimal beam-

former and interference reference nullformer are further processed using the TRINICON algorithm to remove cross-correlations between the two channels.

Chapter 3

Estimated Wiener Filter

3.1 Outline

This chapter describes a simple, novel method for estimating the single channel Wiener filter using two spatially robust beamformers — one directed towards the expected location of the desired source, and the other designed to produce a robust null directed towards the desired signal (the nullformer). The nullformer output is used to provide an instantaneous estimate of the interferer statistics, which can be used to derive a simple estimate of the single channel Wiener filter.

3.2 Estimated Multi-channel Wiener Filter

3.2.1 Introduction

The multichannel Wiener filter (MWF) is often presented as an optimal technique for interference/noise reduction in many areas of signal processing [McCowan and Bourlard, 2003; Van Trees, 2004]. The technique involves designing a set of filters ($\mathbf{w}$) which minimises the mean squared error between the desired signal (s) and the filtered noisy signal received at a sensor array for each wavenumber/frequency k and time index t in the short time-frequency domain, expressed as

$$\mathbf{w}^H[k,t]\mathbf{x}[k,t] = \mathbf{w}^H[k,t]\left(s[k,t]\boldsymbol{\psi}_{\mathrm{s}}[k] + \mathbf{v}[k,t] + \mathbf{n}[k,t]\right) \tag{3.1}$$

where k denotes the wavenumber ($k = 2\pi f/c_0$ — where f is the frequency in Hertz, and c_0 is the speed of sound), $\boldsymbol{\psi}_{\mathrm{s}}$ describes the acoustic transfer function from the desired source location to each of the M microphones in the array, $\mathbf{v}$ represents the interference received at the array, and $\mathbf{n}$ denotes sensor noise. Using the mean squared error minimisation criteria, the Wiener filter solution can be expressed as (2.80):

$$\mathbf{w}[k,t] = \mathbf{R}_x[k,t]^{-1}\mathbf{r}_{\mathbf{sx}}[k,t] \tag{3.2}$$

where $\mathbf{R}_x$ is generated recursively as

$$\mathbf{R}_x[k,t] = \alpha\mathbf{R}_x[k,t-1] + (1-\alpha)\mathbf{x}[k,t]\mathbf{x}^H[k,t] \tag{3.3}$$

and $\mathbf{r}_{\mathbf{sx}}$ is the cross correlation between the array output and the desired signal.

This can be factorised into the well known MVDR (Minimum Variance Distortionless Response) beamformer plus a single channel Wiener filter post-processor, as seen in the previous chapter (2.90) — with the addition of a sensor noise/regularisation matrix $\mathbf{R}_{\mathrm{n}}$.

$$\mathbf{w} = \frac{\sigma_{\mathrm{s}}^2}{\sigma_{\mathrm{s}}^2 + \sigma_{\mathrm{v}}^2} \frac{\mathbf{R}_{\mathrm{v}}^{-1} \boldsymbol{\psi}_{\mathrm{s}}}{\boldsymbol{\psi}_{\mathrm{s}}^H \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1} \boldsymbol{\psi}_{\mathrm{s}}} \tag{3.4}$$

In most practical scenarios, the signal or interference statistics (σ_{s}^2 or σ_{v}^2) are not available and must be estimated. Existing estimation techniques in speech enhancement involve the use of voice activity detection (VAD) to generate speech and interference statistics by detecting pauses during speech. Issues with VAD include false positives/negatives where interference may be falsely detected during a speech utterance or vice versa, the likelihood of which increases as the signal to interference ratio decreases [Catic et al., 2010]. In this chapter a technique is presented for estimating the interference statistics (σ_{v}) and filtering an arbitrary noisy signal, where the approximate position of the desired signal is known and exploited to produce two beamformers, a primary directed at the source, and a secondary designed to suppress sources from a specified region enclosing the assumed source location.

The generalized sidelobe canceller (GSC) [Griffiths and Jim, 1982; Van Trees, 2004] attempts to perform a similar task by constructing two beamformer outputs — the first an LCMV (linearly constrained minimum variance) beamformer directed towards the source; the second an adaptive least mean squares/regularised least squares (LMS/RLS) beamformer derived from an orthogonal blocking matrix, a set of beamformers which null the source signal. The performance of the GSC system depends on the formulation of the blocking matrix [Griffiths and Jim, 1982; Gannot et al., 2001; Gannot and

Cohen, 2002] and the convergence performance (speed and filter accuracy) of the adaptive filter. The blocking matrix formulation achieves its aim of suppressing the desired signal by producing a set of precise null generating beamformers, however these are typically not robust to errors in the desired source position which can lead to desired signal leakage into the adaptive filter, reducing performance.

The Wiener filter estimation method presented in this chapter provides a (near) instantaneous estimate of the interference power spectra which can be used to find an estimate of the signal power, allowing a single channel Wiener filter estimate to be produced quickly without the issues of optimal blocking matrix design and LMS/RLS convergence issues present in GSC based designs.

3.2.2 Filter Estimation

The method can be summarised as designing two beamformers to collect two signals, one of which represents an estimate of the interference in the environment. The first beamformer would use the MVDR algorithm to receive the desired signal (plus residual interference), and the second beamformer is designed to receive only interference, by placing a null directed at the desired signal position.

The MVDR beamformer is widely used in situations where the direction of arrival (far-field) or position (near-field) of the desired source is known. The beamforming weight solution is given as the right hand side of (3.4). In most practical scenarios, the interference correlation matrix $\mathbf{R}_{\mathrm{n}}$ is unknown and the total signal input correlation matrix $\mathbf{R}_x$ is used in its place [Van Trees, 2004; Lorenz and Boyd, 2005; Ba et al., 2007]. The output of this type

of MVDR system is the original desired signal plus attenuated interference. Dropping the frequency/time indices for clarity, the output is

$$y = \mathbf{w}^H \mathbf{x} = s + \frac{\boldsymbol{\psi}_{\mathrm{s}}^H \mathbf{R}_x^{-1} [\mathbf{v} + \mathbf{n}]}{\boldsymbol{\psi}_{\mathrm{s}}^H \mathbf{R}_x^{-1} \boldsymbol{\psi}_{\mathrm{s}}} \tag{3.5}$$

To use the single channel Wiener filter post-processor, an estimate of the interference statistics is required. Our method obtains this by designing a complementary beamformer (or nullformer) to remove the desired signal, leaving behind only interference. The interference estimate is obtained by finding a beamforming solution ($\mathbf{v}$) which directs a null towards the expected location of the desired source. This can be achieved through adaptive and non-adaptive methods by maximising the interference to signal ratio of the output of the nullformer.

3.2.3 Fixed Nullformer

The simplest method of obtaining nullforming weights is to design a non-adaptive system, where the beampattern has a fixed null directed at the desired source and the main lobe directed to maximise the response elsewhere.

Assuming a distribution of interference which can be modelled using the well known isotropic interference correlation function, where the a^{th} row and b^{th} column can be expressed using the distance between the a^{th} and b^{th} elements ($r_{a,b}$) and the wavenumber k from (2.119) as

$$\mathbf{R}_{\mathrm{v}}[a, b] = j_0(k r_{a,b}) \tag{3.6}$$

the nullformer can be designed by optimising the Rayleigh quotient representing the interference to signal (ISR) ratio of the output of the array

$$\mathrm{ISR} = \frac{\mathbf{v}^H \mathbf{R}_{\mathrm{v}} \mathbf{v}}{\mathbf{v}^H \mathbf{R}_{\mathrm{s}} \mathbf{v}} \tag{3.7}$$

A solution can be obtained by considering the equivalent problem of minimising the signal content of the output subject to maintaining the total interference component of the output of the array to some constant level, similar to the maximum SINR solution detailed in Section 2.3.6. The nullformer weights can be obtained through the generalised eigenvalue equation (2.102), and selecting the eigenvector associated with the *smallest* eigenvalue (corresponding to the minimum SINR solution).

3.2.4 Adaptive Nullformer

In most practical scenarios the interferer sources do not remain fixed in both intensity and position relative to the microphone array. As a result the fixed nullformer may not optimally detect moving sources with varying intensity. For this reason, an adaptive nullformer is desirable. This allows the nullformer to track changing interference statistics leading to improved interference estimates for the Wiener filter. The fixed design solution can be converted to an adaptive nullformer design by replacing the assumed fixed inteference correlation matrix $\mathbf{R}_{\mathrm{v}}$ in (2.102) with the continuously updated input correlation matrix $\mathbf{R}_{\mathrm{x}}$, and again selecting the eigenvector solution associated with the smallest eigenvalue as in the previous subsection.

This eigenvector approach to computing an interference estimate produces the optimal nullforming solution to minimise signal received, whereas a full set of orthogonal beamformers as formulated using a GSC-like technique may lead to signal leakage in the blocking matrix path.

An instantaneous estimate for the interference variance at each frequency can be obtained by computing an equalised output of the nullformer. The equalisation is performed by computing the relative expected interference

power ratio of the primary beamformer and nullformer.

$$\hat{\sigma}^2_{\mathrm{v,instant.}} = \|\mathbf{w}^H_{\mathrm{null}}\mathbf{x}\|\sqrt{\left(\frac{\mathbf{w}^H_{\mathrm{MVDR}}\mathbf{R}_{\mathrm{v}}\mathbf{w}_{\mathrm{MVDR}}}{\mathbf{w}^H_{\mathrm{null}}\mathbf{R}_{\mathrm{v}}\mathbf{w}_{\mathrm{null}}}\right)} \tag{3.8}$$

A long term smoothed estimate can be obtained by recursively updating σ^2_{v}.

$$\hat{\sigma}^2_{\mathrm{v}}[t] = \beta\hat{\sigma}^2_{\mathrm{v,instant.}} + (1-\beta)\sigma^2_{\mathrm{v}}[t-1] \tag{3.9}$$

The parameter $0 < \beta < 1$ controls the smoothness of the filter updates, a large value results in a rapidly responding filter estimate, a smaller value a longer term estimate.

The signal power can be estimated from the output power of the primary beamformer and the long term interference estimate. In a manner similar to (3.9), a smoothing parameter can be introduced to prevent the filter estimate from rapidly fluctuating.

$$\sigma^2_{\mathrm{s}}[t] \simeq \sigma^2_x[t] - \hat{\sigma}^2_{\mathrm{v}}[t] \tag{3.10}$$

3.2.5 Desired Source Correlation

The source correlation matrix ($\mathbf{R}_{\mathrm{s}}$) can be modelled by considering the expected location of the desired source. In many practical scenarios, the location of the desired source is approximately known. For example, when designing an array for a cellphone, a reasonable assumption to make would be that the users mouth is located close to the microphones on the bottom of the device. Like the MVDR beamformer, an exact location can be assumed in which case the correlation matrix can be obtained from the acoustic transfer function vector $\boldsymbol{\psi}_{\mathrm{s}}$

$$\mathbf{R}_{\mathrm{s}} = \sigma^2_{\mathrm{s}}\boldsymbol{\psi}_{\mathrm{s}}\boldsymbol{\psi}^H_{\mathrm{s}} \tag{3.11}$$

However this formulation presents a major problem: the matrix $\mathbf{R}_s$ fails to take into account positional error. As nulls tend to be precise in nature, a position mismatch between the expected location and the actual location of the source would lead to errors in the interference spectra estimate for the Wiener filter, degrading the output signal quality. A more robust technique is possible by requiring the source correlation matrix $\mathbf{R}_s$ to represent a distribution of possible desired source locations.

In [Teal et al., 2002b], the authors describe a correlation function for any general distribution of far-field sources by using a spherical harmonic description of plane waves and present a number of results for various types of angular source position distributions. This allows the computation of a source correlation matrix for a far-field scenario which would be useful for distant talker applications. In this chapter it is assumed that the desired source is located close enough to the array to require a near-field treatment incorporating source-array distance information.

3.2.6 Near-field Source Correlation

Of interest to many speech applications and for this work in particular, is the near-field source correlation. The far-field assumption may not be valid for certain scenarios such as hand-held cellphone usage for example, where the distance between the microphone array and the desired source is comparable to or less than the wavelength of sound. In this scheme, the wave fronts from the source to the microphone array are spherical, which is not modelled accurately using far-field assumptions.

The authors in [Dam et al., 2004; Davis et al., 2005] present a beamforming technique in which a probabilistic near-field source distribution is used to

generate a source correlation matrix used to help design the beamforming solution. In this paper, the correlation matrix is applied to the problem of producing a robust null rather than a robust primary beamformer — since it is assumed that the MVDR solution for a compact array exhibits intrinsic robustness to positional mismatch.

For near-field sources the variation in source distance needs to be considered in addition to angular position. In 3D using the spherical coordinate system, the correlation function between the i^{th} and j^{th} microphones can be described as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \int_r \int_\theta \int_\phi \rho(r,\theta,\phi)\, \psi_a \psi_b^*\, r^2\, dr\, d\Omega, \tag{3.12}$$

where the ψ terms represent the near-field acoustic transfer function. For example, the point source transfer function can be decomposed in terms of spherical harmonics as [Colton and Kress, 1998]

$$\psi_a = ik \sum_{n=0}^{\infty} \sum_{m=-n}^{n} j_n(kr) h_n(kr_a) Y_n^m(\Omega_a) Y_n^{m*}(\Omega_a) \tag{3.13}$$

where r denotes the source radius, r_a the microphone radius, j_n denotes the spherical Bessel function of order n, h_n denotes the spherical Hankel function of order n, the $Y_n^m(\Omega)$ terms denote the spherical harmonic functions for a given angular position, and $\Omega = (\theta, \phi)$.

The integral (3.12) has no simple solution for a Gaussian-like distribution of source locations. The correlation function can be approximated through numerical integration by substituting in the appropriate source position distribution function. In this chapter, the position distribution is assumed to be a spherically symmetric Gaussian distribution centred close to the microphone array. In Section 4.3.3, an analytic approximation for (3.12) is developed and could be used as an alternative to numerical integration.

3.2.7 Simulation Setup

Simulations were conducted to compare the estimated Wiener filter technique with solely MVDR beamforming and a perfect multichannel Wiener filter derived from perfect signal and interference information. The estimated Wiener filter was evaluated by placing the desired source at a fixed location 30cm from the centre of the microphone array and the interferers evenly distributed 1m from the centre of the array. The microphone array used was a 4 omnidirectional element circular array with a radius of 1cm. The audio was sampled at 8kHz and the plane-wave (far-field) and point-source (near-field) models of sound used to generate the acoustic transfer function vectors ($\boldsymbol{\psi}$). The sources were placed in a lightly reverberant 3D room to simulate diffuse interference, which was generated using the image source method [Allen and Berkley, 1979]. The wall reflection coefficients were set to 0.3 and up to 4th order reflections were generated. The FFT block size used for computing the MVDR filters was set to 128. The target signal to interference ratio was set to 0 dB.

The near-field correlation matrix was derived by assuming a 3D Gaussian distribution of sources centred on (0.30m, 0m, 0m) with a variance of 5cm. This represented a potentially moving desired source which was predominantly located within 2 standard deviations of the centre, a model chosen to represent head movement relative to some fixed array location. The nullformer weights were derived by inserting a regularisation matrix into the source correlation matrix in order to improve robustness at low frequencies. The sensor noise correlation matrix was set as $\mathbf{R}_{\mathrm{n}} = 10^{-8} \times \mathbf{I}$, where $\mathbf{I}$ is the $M \times M$ identity matrix.

3.2.8 Results

The performance of the estimated Wiener filter was evaluated for both far-field and near-field speech sources using three criteria: SINR, signal distortion and perceptual quality (using the ITU P.862 PESQ standard [ITU-T, 2002]). The estimated filter was compared with the ideal filter, derived from perfect signal and interference knowledge. This represented the best case scenario for a voice activity detector based system with zero false positives/negatives.

The estimated filter produced an output which closely matched the ideal filter during speech utterances with an average improvement in SINR of close to 17dB (5dB on top of the MVDR beamformer) for the near-field simulation (Figure 3.1). The estimated filter shows a slight apparent improvement over the ideal filter in some regions of audio, however this is due to imperfections in the filter design resulting in errors in the signal and interference spectra estimates, which leads to the filter aggressively removing both speech and interference in some frequency bands resulting in a higher SINR. The side effect of this aggressive filtering is an increase in signal distortion during the speech utterances relative to the perfect filter (Figure 3.2), which degrades the relative signal quality. The aggressive filtering arises primarily from conditioning problems in the solution for the beamforming weights. The solution to the eigenvalue problem (2.102) remains ill-conditioned at low frequencies, requiring regularisation (increasing the value of the diagonal entries of $\mathbf{R}_{\mathrm{n}}$) for a numerically stable solution. This results in reduced accuracy in the interference spectra estimate for the Wiener filter. Despite these limitations, the estimated filter performs well for the task of improving speech intelligibility where the simulations have shown a significant improvement in perceptual

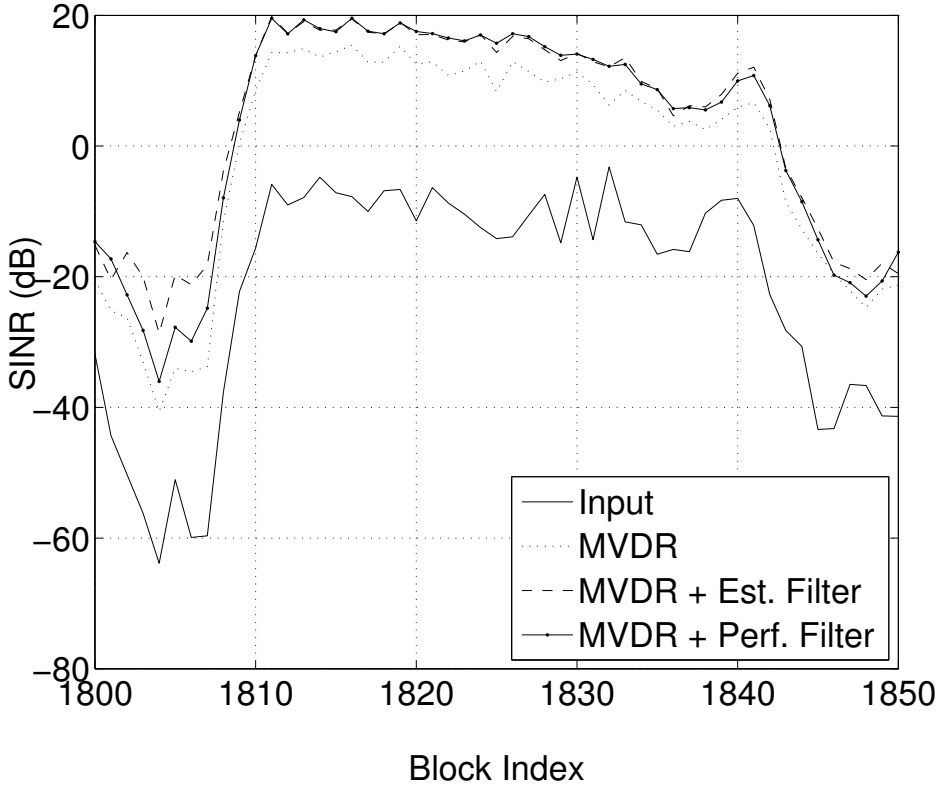

Figure 3.1: SINR of a short utterance of speech for the near-field method comparing the input, MVDR, MVDR plus estimated filter and MVDR plus perfect filter

quality as evaluated using the PESQ standard (Table 3.1).

3.2.9 Discussion

The main deficiency of the proposed method against the traditional VAD based methods is the inability to filter out interferers which lie in the direction of the spatial null created by the nullformer. In most practical implementations,

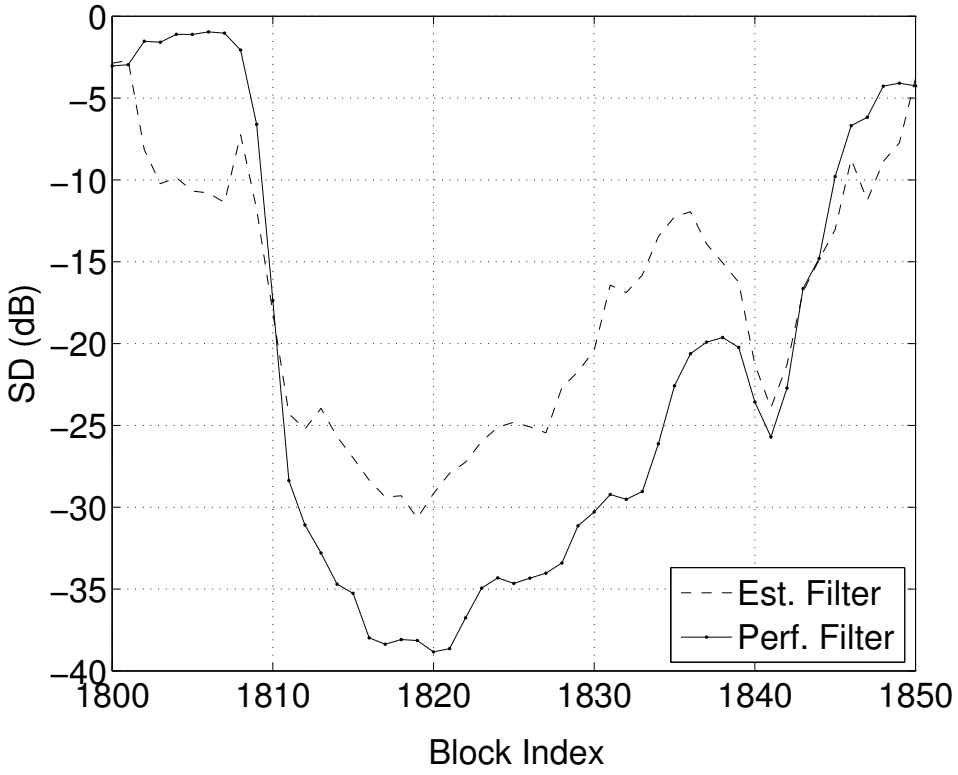

Figure 3.2: Signal distortion measures for a short utterance of speech for the near-field method comparing the MVDR plus estimated filter and the MVDR plus perfect filter

	Score
Input	1.1
MVDR	1.7
MVDR + Est. Filter	2.2
MVDR + Perf. Filter	2.7

Table 3.1: PESQ scores

the limited number of microphones and aperture width will result in this problem as the null size is restricted by these parameters. However, unlike VAD based methods, this technique does not require detection of speech which can be difficult in low SINR environments. This makes the proposed approach well suited for many practical scenarios. Additionally, although this has been presented as a speech enhancing method, no assumptions other than the location of the desired signal have been made, making it well suited for other signal filtering operations where the position of the desired source is known. For the specific example of speech enhancement, the imperfect estimated filter performs reasonably well against the best case scenario where the Wiener filter can be perfectly estimated.

Chapter 4

Spatially Robust Beamforming

4.1 Outline

This chapter covers beamformer spatial robustness; it is assumed that the desired source location can vary with some angle (far-field) or position (near-field) from the given location. The far-field section focuses on using the von Mises(-Fisher) distribution to model angle variance to design robust beam/nullformers. The near-field section presents two formulations to model position variance — a uniform probability distribution (equal probability within a volume), and a radial Gaussian model.

4.2 Far-field Beamforming

4.2.1 Introduction

An issue of interest in the field of beamforming is spatial robustness. Beamforming solutions such as the minimum variance distortionless response (MVDR) and its related algorithms are often presented in the literature. In these methods, the signal to interference plus noise (SINR) ratio is optimised given an exact source location and precise interference/noise statistics.

In some situations the actual location of a desired source may not match the assumed location used to derive the beamformer solutions. This may lead to reduced SINR gain when a mismatch occurs. Existing approaches to this problem include target tracking, where voice activity detection (VAD)/signal detection and non-stationarity assumption methods are commonly used to train the steering vectors and/or interference and noise statistics [Gannot et al., 2001; Gannot and Cohen, 2002].

Previous approaches to the spatial robustness problem include generalised eigenvalue (GEVD) based beamformers [Shahbazpanahi et al., 2003; Dam et al., 2004], which can be used to design solutions optimised for a particular region of interest. In these methods, the correlation functions are computed by integrating the correlation between two microphones for a region of possible desired source directions.

Additionally, it may be desirable to design blocking beamformers (nullformers) which suppress signals originating from a particular direction. Typical null designs such as those used in LCMV and GSC beamforming [Van Trees, 2004] result in very narrow suppression angular regions which are sensitive to direction mismatch between the actual and estimated direction of arrival.

For example the generalised sidelobe canceller (GSC) algorithm relies on a set of orthogonal blocking beamformers (suppressing the desired source) to control adaptation. If a misalignment occurs, there can be significant desired signal leakage into the blocking path, leading to reduced performance. In a manner similar to many implementations of the MVDR beamformer, training methods such as VAD can be used to correct the steering vectors.

VAD-based methods used to adapt the steering vectors and/or interference and noise statistics are not robust in noisy environments [Davis et al., 2005; Catic et al., 2010]. In low SINR conditions, the probability of false positives/negatives increases which leads to incorrect steering vector and/or interference/noise estimation.

In this section, a method for designing robust microphone beamformers is presented based on the use of distribution modelling of the desired source direction of arrival. A simple analytic result for the spatial correlation function due to a distribution is used to design the beamformers — a technique which greatly reduces the computational complexity of the beamformer weight solutions.

4.2.2 Robust Maximum Eigenvalue Beamforming

A method of introducing directional robustness into deriving the beamforming weights is rather than assuming a single steering direction, to design the weights to account for multiple possible steering directions. The maximum SINR beamformer solution (2.102) introduced in Section 2.3.6 provides a framework for designing robust beamformers.

The desired source spatial correlation matrix can be derived by considering

multiple possible steering vectors weighted by some probability.

$$\mathbf{R}_{\mathrm{s}} = \frac{1}{N} \sum_{n=1}^{N} p_n \boldsymbol{\psi}_n \boldsymbol{\psi}_n^H \tag{4.1}$$

where p_n denotes the probability weighting for the n^{th} source, and $\boldsymbol{\psi}_n$ denotes the transfer function vector from the source to the microphone.

This can be extended to a far-field continuous distribution in a 2D plane by integrating over a circle. The entries of the $\mathbf{R}_{\mathrm{s}}$ matrix can be expressed as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \int_{\phi} p(\phi,\phi_0)\psi_a(\phi,\phi_a)\psi_b^*(\phi,\phi_b)\,d\phi \tag{4.2}$$

where $p(\phi,\phi_0)$ denotes the probability that the desired source is located in the direction ϕ given a central angle ϕ_0, ψ_a denotes the steering coefficient from the desired source to the a^{th} microphone (and similarly for the b^{th} microphone). Similarly a far-field 3D description is expressed as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \iint_{\Omega} p(\Omega,\Omega_0)\psi_a(\Omega,\Omega_a)\psi_b^*(\Omega,\Omega_b)\,d\Omega \tag{4.3}$$

where the Ω vectors are shorthand descriptions of the inclination and azimuth angles (θ,ϕ), and $d\Omega = \sin\theta\,d\theta\,d\phi$.

In general the interference spatial correlation matrix is unknown for many scenarios. A reasonable assumption in reverberant environments is to assume that the interference is isotropic in nature [Ward and Elko, 1997]. The 2D or 3D isotropic interference spatial correlation functions [Teal et al., 2002b] can be used to provide an estimate for the correlation matrix. The 2D function can be expressed as

$$\mathbf{R}_{\mathrm{v}}[a,b] = J_0(kd_{ab}) \tag{4.4}$$

where J_0 denotes the zeroth order cylindrical Bessel function, k the wavenumber, and d_{ab} the distance between the a^{th} and b^{th} microphones in the array.

The 3D isotropic function is similarly (from Section 2.3.7)

$$\mathbf{R}_{\mathrm{v}}[a,b] = j_0(kd_{ab}) \tag{4.5}$$

where j_0 denotes the zeroth order spherical Bessel function.

In situations where more knowledge on the environment is available, alternative interference correlation matrices can be substituted.

The integral formulation for the desired source spatial correlation function poses a computational complexity problem when designing the beamformer. The integral typically does not have a simple analytic solution, meaning costly numerical integration techniques are required to compute the correlation matrix. However, certain probability distribution functions exhibit mathematical properties which can lead to analytical solutions to the integrals in (4.2 and 4.3).

4.2.3 Robust Nullforming

In some applications it may be desirable to suppress rather than enhance an source arriving from some direction. Again we assume that the target source lies within an uncertain region. For the robust beamformer the optimisation criterion for designing weights was defined as finding weights which maximised the SINR. In robust null design the optimisation criteria can be defined as doing the opposite, i.e., maximising the interference to signal ratio. The GEVD equation (2.102) to derive the nullformer weights is similar to the beamformer, with the desired source and interference correlation matrices exchanged. The nullformer solution can be obtained by finding the eigenvector ($\mathbf{w}$) associated with largest eigenvalue (λ) in the equation

$$\left[\mathbf{R}_{\mathrm{s}} + \sigma_{\mathrm{n}}^2\mathbf{I}\right]\mathbf{w} = \lambda\mathbf{R}_{\mathrm{v}}\mathbf{w} \tag{4.6}$$

As in the robust beamformer case, this formulation allows for a spatially robust solution depending on the design of the target source correlation matrix $\mathbf{R}_{\mathrm{s}}$.

4.2.4 von Mises Distribution based Beamformer

Two Dimensional Modelling

The correlation between two microphones due to a plane-wave source originating from angle ϕ can be expressed as

$$\psi_a(\phi, \phi_a)\psi_b^*(\phi, \phi_b) = e^{i\mathbf{k}\cdot\mathbf{r}_a}e^{-i\mathbf{k}\cdot\mathbf{r}_b} = e^{i\mathbf{k}\cdot\mathbf{r}_{ab}} \tag{4.7}$$

where $\mathbf{k} = k\,[\cos\phi\ \sin\phi]^T$ denotes the wavevector describing the wave originating from a source direction ϕ, $\mathbf{r}_a = r_a\,[\cos\phi_a\ \sin\phi_a]^T$ denotes the microphone position vector, and $\mathbf{r}_{ab} = \mathbf{r}_a - \mathbf{r}_b$ denotes the vector between the two microphones. In terms of Bessel functions, (4.7) can be expressed as

$$\psi_a(\phi, \phi_a)\psi_b^*(\phi, \phi_b) = \sum_{n=-\infty}^{\infty} i^n J_n(kr_{ab})e^{in(\phi_{ab}-\phi)} \tag{4.8}$$

The von Mises distribution [Teal et al., 2002a] provides a suitable model for describing the variation in direction of arrival of a sound source. The von Mises density function is

$$p(\phi, \phi_0) = \frac{e^{\kappa\cos(\phi-\phi_0)}}{2\pi I_0(\kappa)} \tag{4.9}$$

where κ describes the shape of the distribution, and is analogous to the parameter σ of the normal distribution, ϕ_0 denotes the centre of the distribution — the look direction of the beamformer in this application, and I_0 is the zeroth-order modified cylindrical Bessel function.

The authors of [Teal et al., 2002a] derive the correlation function result as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{1}{I_0(\kappa)} \sum_{n=-\infty}^{\infty} i^n J_n(kr_{ab}) I_n(\kappa) e^{in(\phi_{ab}-\phi_0)} \tag{4.10}$$

In Section 4.2.9, (4.10) is derived and it is demonstrated that it can be simplified to a simple novel ratio function by exploiting the properties of the modified Bessel functions.

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{J_0(z)}{I_0(\kappa)} \tag{4.11}$$

where

$$z = \sqrt{(kr_{ab})^2 - \kappa^2 - 2i\kappa kr_{ab}\cos(\phi_{ab}-\phi_0)} \tag{4.12}$$

is a complex number which is related to the non-isotropy of the sound field.

If κ is zero, the correlation function is that of an isotropic field — the $(kr_{ab})^2$ term dominates. When κ becomes large, the field becomes more directional — the κ^2 term dominates. This behaviour can be parametrised by considering the ratio

$$\eta = \frac{(kr_{ab})^2}{\|z\|^2} \tag{4.13}$$

which is equal to 1 if the field is isotropic ($\kappa = 0$) and approaches zero as κ tends to infinity.

The distribution shape for varying values of κ is demonstrated in Figure 4.1. Smaller values of κ correspond to a broader distribution, larger values correspond to a more compact localised distribution. For the uniform distribution case ($\kappa = 0$), the correlation function becomes the 2D isotropic interference correlation function in (4.4).

Using this correlation function, it is possible to design spatially robust beamformers/nullformers using the generalised eigenvalue solution, described

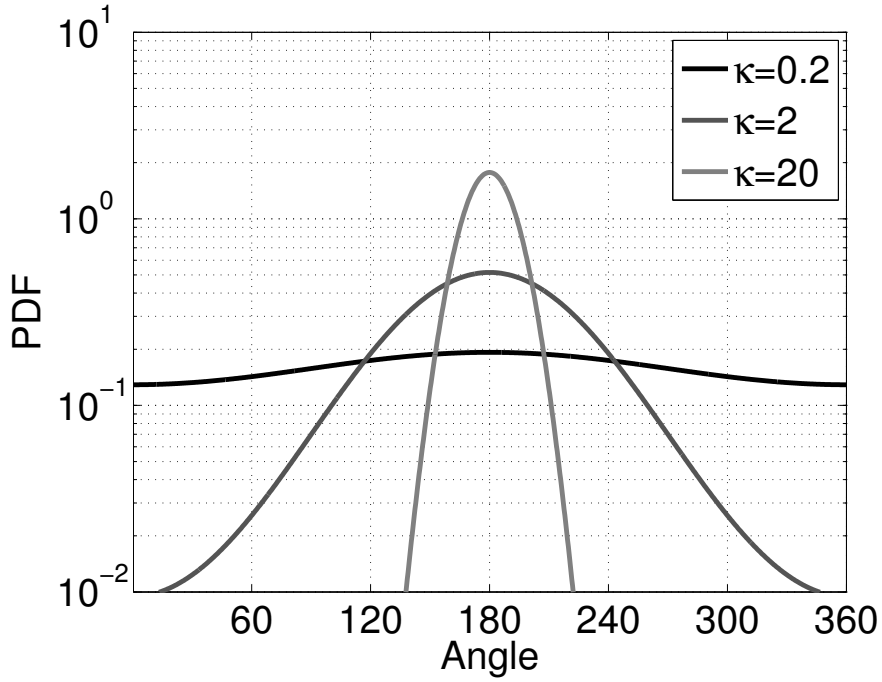

Figure 4.1: The von Mises probability density function centred at $\phi_0 = 180°$ for varying values of κ.

in Sections (4.2.2 and 4.2.3), by inserting the appropriate spatial correlation matrices.

Three Dimensional Modelling

In [Mammasis and Stewart, 2010], the authors describe a similar correlation function for the 3D von Mises-Fisher distribution. Like the 2D case, the correlation function can be computed using a series solution for the integral in (4.3).

The 3D von Mises-Fisher distribution is defined as [Fisher, 1953]

$$p(\Omega, \Omega_0) = \frac{\kappa}{4\pi \sinh \kappa} e^{\kappa[\cos\theta \cos\theta_0 + \sin\theta \sin\theta_0 \cos(\phi - \phi_0)]} \tag{4.14}$$

In [Mammasis and Stewart, 2010] the authors derive the correlation function using the spherical harmonic description of plane waves as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{4\pi}{i_0(\kappa)} \sum_{n=0}^{\infty} \sum_{m=-n}^{\infty} i^n j_n(kr_{ab}) \iota_n(\kappa) Y_n^m(\Omega_{ab}) Y_n^{m*}(\Omega_0) \tag{4.15}$$

where ι_n denotes the n^{th} order modified spherical Bessel function, Y_n^m denotes the (n,m) order spherical harmonic function, Ω_{ab} denotes the solid angle between the a^{th} and b^{th} microphones, and Ω_0 denotes the target direction.

In Section 4.2.10 the result in (4.15) is derived and it is shown that a further simplification is possible by exploiting the properties of the (modified) spherical Bessel functions, analogous to those of the cylindrical Bessel functions. The result is a novel simple ratio function involving only the zeroth order (modified) spherical Bessel functions.

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{j_0(z)}{\iota_0(\kappa)} \tag{4.16}$$

where

$$z = \sqrt{(kr_{ab})^2 - \kappa^2 - 2i\kappa kr_{ab}\cos(\Psi)} \tag{4.17}$$

is a complex quantity relating to the (non)isotropy of the sound field similar to (4.12), and

$$\cos\Psi = \cos\theta_{ab}\cos\theta_0 + \sin\theta_{ab}\sin\theta_0\cos(\phi_{ab} - \phi_0) \tag{4.18}$$

is the cosine of the angle between the centre of the source distribution and the vector connecting the a^{th} and b^{th} microphones.

Analogous to the 2D case, the uniform distribution ($\kappa = 0$) results in the well known 3D isotropic interference correlation function in (4.5).

4.2.5 Results

The robust method was compared with the existing MVDR solution by generating a set of beamformers using different values of κ corresponding to wide (small values) or narrow (large values) main lobes. The theoretical SINR improvement (due to isotropic noise) and white noise gain were computed as performance measures. Three frequencies were compared for each microphone array layout — 550Hz, 1.1kHz and 2.6kHz. Three arrays were designed — a compact 4 microphone array, a small teleconferencing array, and a large teleconferencing array.

Compact Array

Using a distribution based approach to designing the beamforming weights results in broad main lobes in the resulting array response beam-patterns if a small value of κ is selected (corresponding to a broad distribution). The compact array with few microphones exhibited a slight broadening when using low values of κ compared with the MVDR solution, as seen in Figure 4.2. The main lobe width, defined in terms of a greater than 0 dB SINR gain — where the signals originating from this region are enhanced relative to the background, increased by up to 15.2° (Table 4.1) when using $\kappa = 0.2$. As the parameter κ shrinks, the SINR gain of the background noise decreases substantially (as much as 10 dB) in exchange for a slight decrease in SINR gain for sources originating from the expected location.

At higher frequencies, the array response is similarly broad, due to the limited aperture and number of microphones. The distribution-based approach still results in a similar level of improvement in spatial robustness, however

Table 4.1: Compact array main-lobe width increase in radians (degrees) compared with MVDR using $\kappa = 0.2$

Freq.	MVDR	Robust ($\kappa = 0.2$)	Width Increase
550Hz	1.641 (94.0)	1.877 (107.5)	0.236 rad. (13.5°)
1100Hz	1.641 (94.0)	1.871 (107.2)	0.230 rad. (13.2°)
2600Hz	1.651 (94.6)	1.917 (109.8)	0.266 rad. (15.2°)

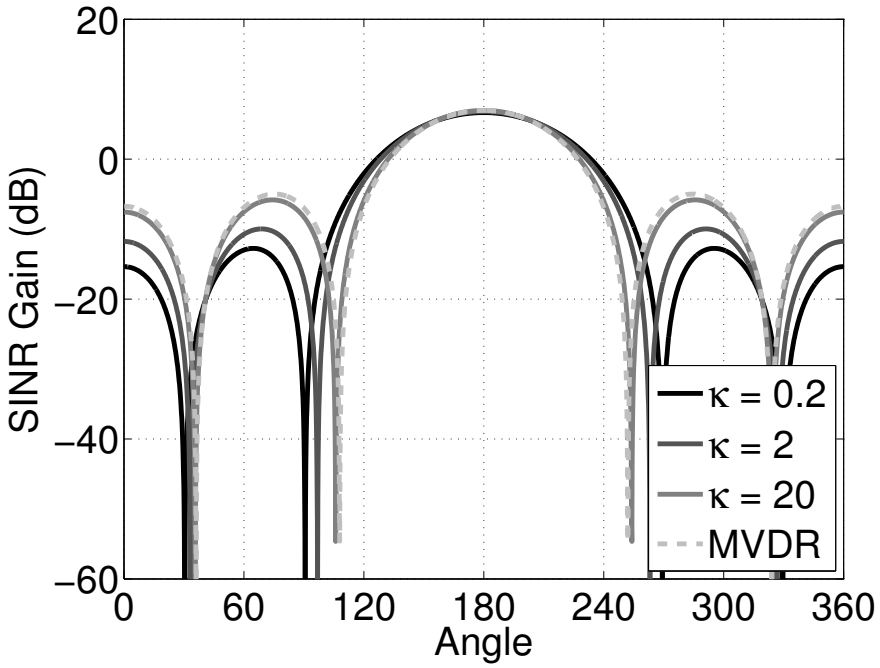

Figure 4.2: Theoretical SINR improvements at 1.1kHz using different values of κ for a 4-element 2cm radius microphone array.

for this particular array size this is unlikely to be necessary as the MVDR solution appears to be spatially robust. The distribution-based approach does lead to improved noise rejection outside the main-lobe which may be advantageous in some scenarios.

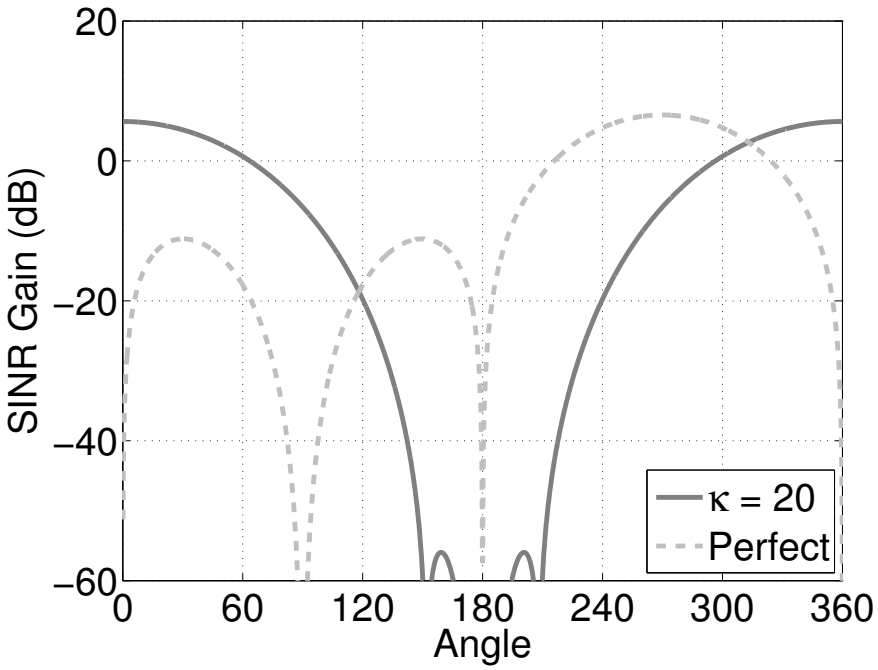

Figure 4.3: Nullformer performance at 1.1kHz for the 4-element 2cm radius array.

The nullformer designed with the robust formulation shows an immediate advantage over the precise null, seen in Figure 4.3. The robust null is able to attenuate signals from the target direction of arrival by at least 50 dB over a angle range of roughly 60°, compared with the precise null produced using the classical design.

Larger Teleconferencing Arrays

Using more microphones and larger apertures results in narrower main lobes when using the MVDR beamformer. For these arrays it becomes more critical that the look direction is accurate for the MVDR solution to operate well. Two larger arrays were tested, the first having 8-elements, 5 cm radius; the

Table 4.2: Small teleconferencing array main-lobe width increase in radians (degrees) compared with MVDR using $\kappa = 0.2$

Freq.	MVDR	Robust ($\kappa = 0.2$)	Width Increase
550Hz	1.173 (67.2)	1.503 (86.1)	0.330 rad. (18.9°)
1100Hz	1.023 (58.6)	1.259 (72.1)	0.236 rad. (13.5°)
2600Hz	1.037 (59.4)	1.289 (73.8)	0.252 rad. (14.4°)

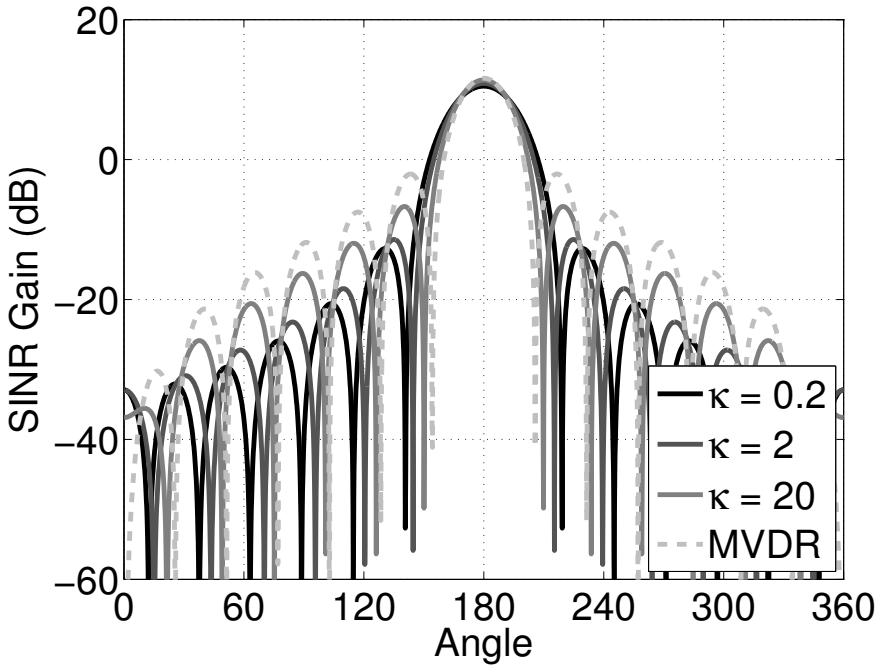

Figure 4.4: Theoretical SINR gain for the 32-element, 10 cm radius array at 1.1kHz.

second 32-elements, 10 cm radius.

For the 5 cm radius array, the robust formulation improves the robustness by a similar amount to the compact array, increasing the main-lobe width by between 13.5 and 18.9° (Table 4.2).

Table 4.3: Large teleconferencing array main-lobe width increase in radians (degrees) compared with MVDR using $\kappa = 0.2$

Freq.	MVDR	Robust ($\kappa = 0.2$)	Width Increase
550Hz	0.801 (45.9)	1.075 (61.6)	0.274 rad. (15.7°)
1100Hz	0.601 (34.4)	0.811 (46.4)	0.210 rad. (12.0°)
2600Hz	0.595 (34.1)	0.877 (50.3)	0.282 rad. (16.2°)

Similar gains in main-lobe width are achieved with the larger array, with a typical improvement of at least 0.2 radians (Table 4.3, Figure 4.4).

Three Dimensional Array

A simple example of a three dimensional system is presented in Figure 4.5, where a 25-element 5 cm radius array is beamformed to a target originating from the negative y direction (0, -1, 0). The robust formulation shows a larger angular range of SINR improvement compared with the MVDR response. Like the 2D case, the side-lobes decrease in intensity (although this is not easily visible in the figure), indicating greater off-target noise rejection compared with the MVDR solution.

Numerical Robustness

As noted in Section 2.3.9, white noise gain (WNG) is an important measure of the ability of the microphone array to tolerate intrinsic errors: sensor noise, calibration errors and so on. In these simulations, σ_{n}^2 was set to 10^{-6}, modelling a 0.1% error in the beamformer weight vectors.

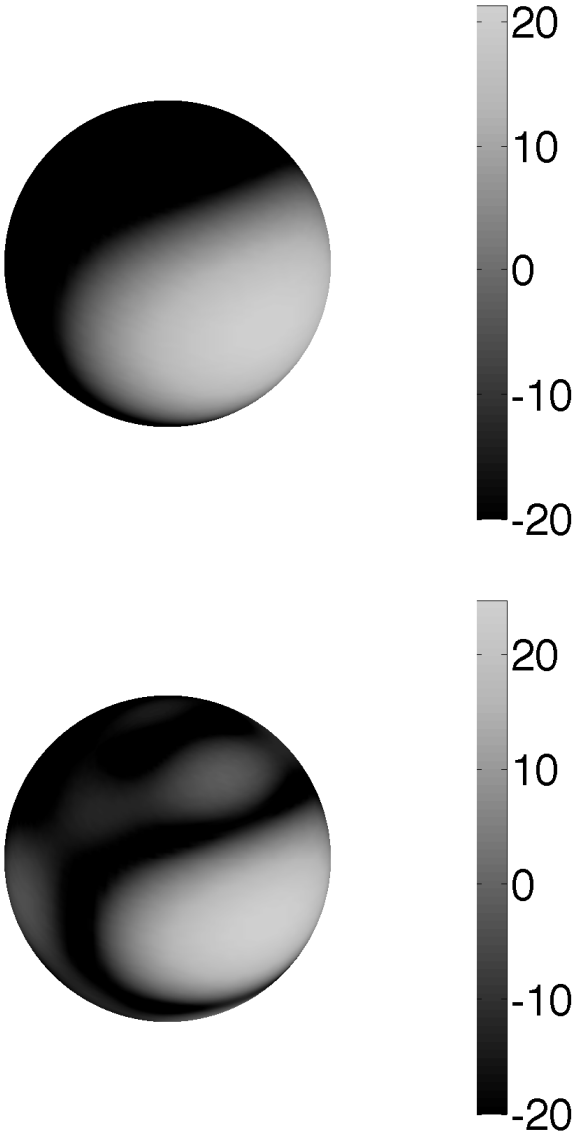

Figure 4.5: SINR gain (in dB) for the a) 3D von Mises-Fisher beamformer ($\kappa = 0.2$) and b) 3D MVDR beamformer at 1.1kHz, with a target direction of (0, -1, 0).

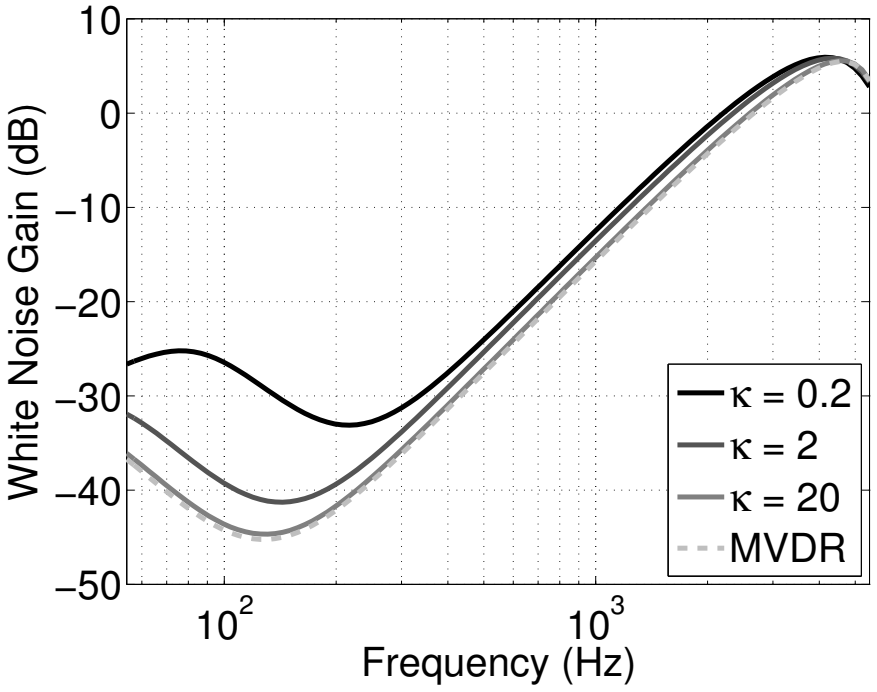

Figure 4.6: White Noise Gain against frequency for the compact array beamformers designed using MVDR and the robust formulation for the 4-element 2 cm radius array.

For the nullformer designs, the white noise gain parameter is redefined as

$$\mathrm{WNG}_{\mathrm{null}} = \frac{\mathbf{w}^H \mathbf{R}_{\mathrm{v}} \mathbf{w}}{\mathbf{w}^H \mathbf{w}} \tag{4.19}$$

Similarly, nullformer numerical robustness is improved by increasing σ_{n}^2.

The robust formulation, seen in Figures 4.6 and 4.7, exhibits improved white noise gain at low frequencies, by as much as 20 dB when using a regularisation parameter of 10^{-6}. This indicates that the spatially robust beamformer is capable of tolerating greater errors in microphone mismatch (array calibration errors and/or intrinsic microphone properties) at low frequencies than the MVDR beamformer. However the robust formulation, like MVDR, does

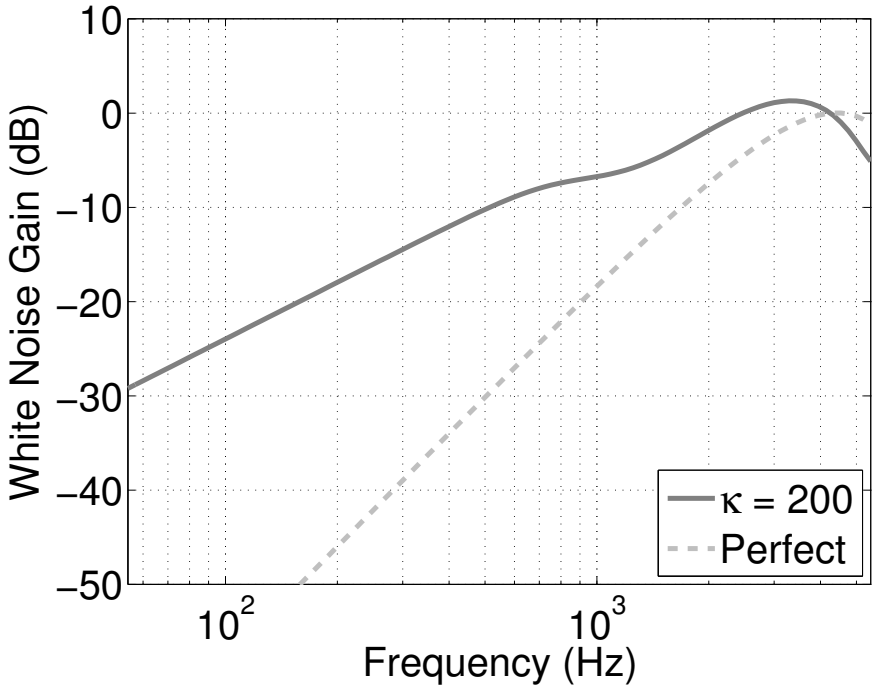

Figure 4.7: White Noise Gain against frequency for the compact array null-formers designed using perfect direction knowledge and the robust formulation for the 4-element 2 cm radius array.

still have poor white noise gain at low frequencies, indicating that significantly more regularisation of the sensor noise correlation matrix (2.54) is required to ensure robustness to intrinsic microphone/array position/mismatch errors.

In Figure 4.8 a simple demonstration of the effect of error on the beamformer response is presented for a low frequency scenario where numerical robustness could be problematic. In this simulation, a 0.1% error was introduced into the weight solution vector $\mathbf{w}$, simulating sensor noise. The noisy sensors exhibit degraded performance as expected, which results in reduced background interference suppression, particularly when using the MVDR

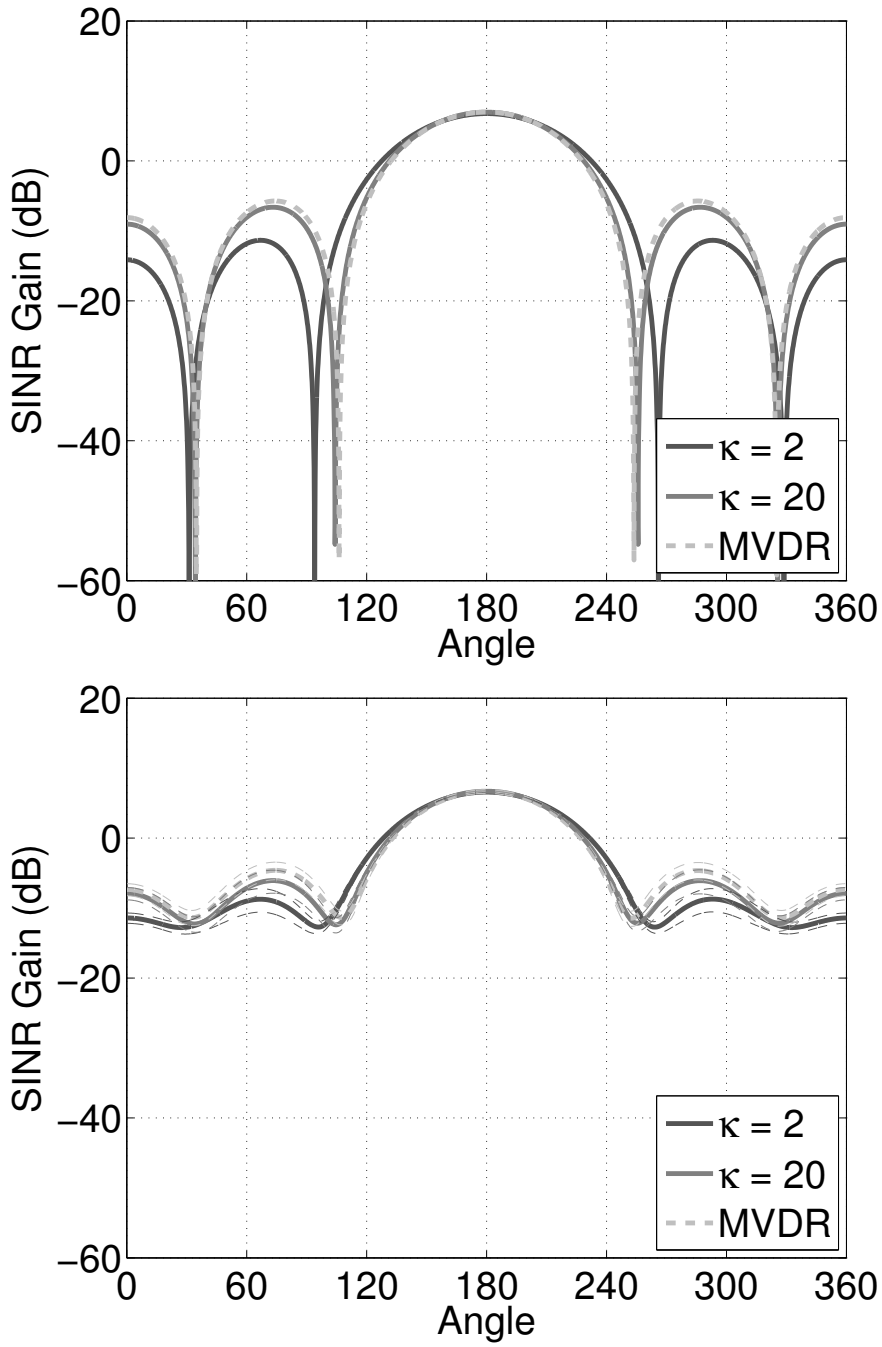

Figure 4.8: Beamformer SINR gain with noiseless sensors (top), and noisy (bottom). 4-element, 2 cm radius array at 250Hz. 95% confidence intervals are displayed for the noisy results.

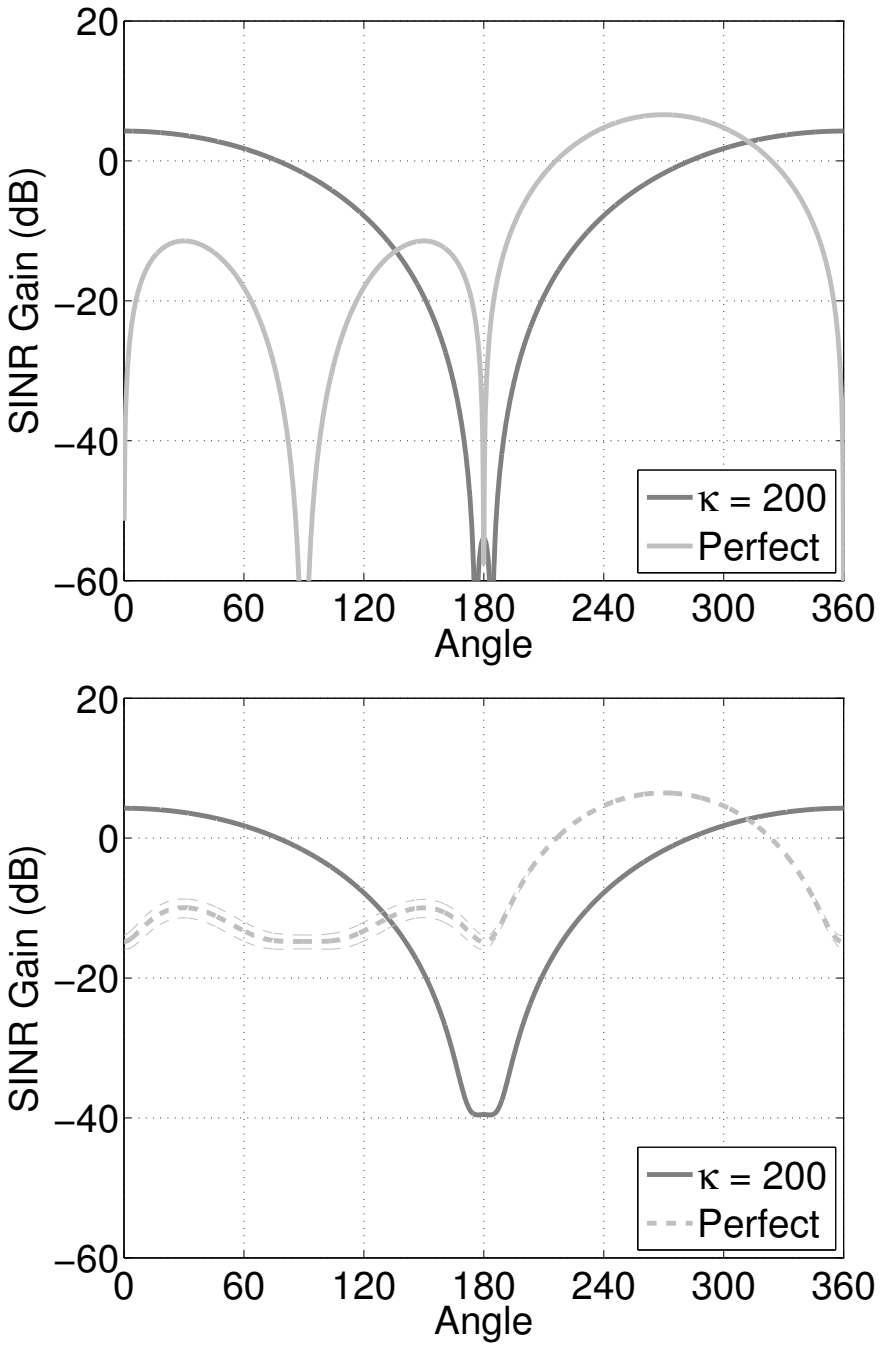

Figure 4.9: Nullformer SINR gain with noiseless sensors (top), and noisy (bottom). 4-element, 2 cm radius array at 250Hz. 95% confidence intervals are displayed for the noisy results.

solution in these simulations. Using the same assumed noise parameters in the beamformer design ($\sigma_{\mathrm{n}}^2 = 10^{-6}$), the von Mises correlation derived beamformer performs significantly better when a broad distribution parameter is chosen. Over 1000 trial runs, in which the weight vector $\mathbf{w}$ was perturbed by a small error, the von Mises derived beamformer showed on average better SINR gain when directed at the target source. Additionally, the background interference is suppressed significantly better (up to 5 dB) compared with the MVDR beamformer.

A simple nullformer example is presented in Figure 4.9, where the weights have been distorted by a small (0.1%) random perturbation. The spatially robust formulation still produces a broad null near the expected direction of arrival (originating from 180°), however it exhibits reduced signal suppression, rising from -50 dB to approximately -40 dB. The result of this is an increase in signal leakage which in some applications may be undesirable. The spatially robust solution still shows a huge advantage over the precise null solution, which shows poor suppression of the target signal at this particular frequency due to its poor white noise gain characteristics.

4.2.6 Application: Simple Adaptive Filtering vs. GSC

An application of the nullforming technique is the implementation of an adaptive filtering system similar to the GSC beamformer (using the time-domain equivalent of (2.71)). For this application, the objective is to emulate the GSC structure by replacing the blocking matrix (multiple perfect nullformers) with a single robust nullformer. The performance was evaluated by placing two non-stationary sources (speech and music, sampled at 44.1 kHz) opposite each other 1 m away from a 4-element, 1 cm radius circular array. The mean

input SINR was set to 0 dB. Adaptation of the simple and GSC methods was achieved using 1024-tap time-domain LMS filters and a step-size parameter $\mu = 10^{-4}$ (found through trial and error to maximise SINR). The 1024-tap fixed-MVDR and fixed-robust beamformers were designed using (2.60 and 2.102) with a 2D diffuse interference correlation matrix, representing a worst-case scenario where no knowledge of the interferer(s) is assumed. The desired source spatial correlation matrix for the robust beamformers was designed using (4.11) using the parameter $\kappa = 700$.

In Table 4.4 the SINR after processing with the simple adaptive filter, GSC, and only fixed-robust/MVDR beamforming is presented. The simple method outperformed the GSC beamformer with an increasing margin (2 to 3 dB) as the location error increased. It was found that the theoretically perfect nulls, designed by constructing vectors orthogonal to the expected transfer function vector, did not perfectly cancel the desired signal even when no location mismatch occurred. This was due to minor arithmetic rounding errors arising from the eigenvalue decomposition method used to derive the blocking matrix, and the corresponding poor white noise gain of the nullformers. This is not unexpected as it was seen in Figure 4.9, that the introduction of a small amount of random noise resulted in almost complete failure of the perfect null design.

4.2.7 Discussion

As κ tends to large values, the resulting beamformer approaches the same performance as the MVDR beamformer. This results from the limiting behaviour of the modified Bessel functions in (4.10) and (4.15). In the limit as κ approaches infinity, the modified cylindrical/spherical Bessel functions

Table 4.4: SINR comparison between the simple adaptive filter system, GSC, fixed robust, and fixed-MVDR beamformers. Input SINR was set to 0 dB.

Mismatch	AF (dB)	GSC (dB)	Fixed-Beam (dB)	Fixed-MVDR (dB)
0°	11.2	9.5	9.7	9.5
5.7° (0.1 rad.)	11.1	9.2	9.6	9.4
11.5° (0.2 rad.)	10.6	8.5	9.2	8.9
17.2° (0.3 rad.)	9.8	7.4	8.4	8.2
22.3° (0.4 rad.)	8.7	6.0	7.3	7.1
28.6° (0.5 rad.)	7.1	4.2	5.7	5.5

approach [Abramowitz and Stegun, 1964]

$$I_n(\kappa) \approx \frac{e^{\kappa}}{\sqrt{2\pi\kappa}} \tag{4.20}$$

and the ratio in the summations in (4.10) and (4.15)

$$\frac{I_n(\kappa)}{I_0(\kappa)}, \frac{\iota_n(\kappa)}{\iota_0(\kappa)} \tag{4.21}$$

approaches 1. In this case the summations in (4.10) and (4.15) simplify to the correlation function due to a single source at a specific angle ϕ_0/Ω_0 and this case the correlation matrix $\mathbf{R}_{\mathrm{s}}$ can be derived as the outer-product of the vector describing the transfer function from the source to each of the microphones ($\boldsymbol{\psi}_{\mathrm{s}}$ in (2.46)).

$$\mathbf{R}_{\mathrm{s}} = \sigma_{\mathrm{s}}^2 \boldsymbol{\psi}_{\mathrm{s}} \boldsymbol{\psi}_{\mathrm{s}}^H \tag{4.22}$$

The GEVD beamformer solution (from (2.102), again neglecting σ_{s} for simplicity) is

$$(\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}})\,\mathbf{w} = \lambda \boldsymbol{\psi}_{\mathrm{s}} \boldsymbol{\psi}_{\mathrm{s}}^H \mathbf{w} \tag{4.23}$$

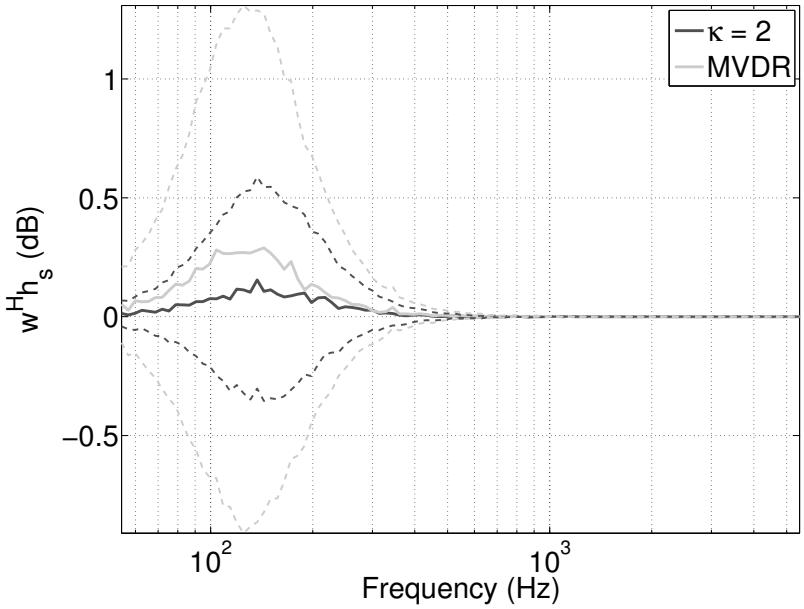

Figure 4.10: Response to a signal located at the mean angle of the source distribution in the presence of sensor noise. A distortionless response corresponds to 0 dB. 95% confidence intervals are displayed as the dashed curves.

Noting that $\boldsymbol{\psi}_{\mathrm{s}}^H \mathbf{w}$ is 1 to satisfy the distortionless constraint, the GEVD equation reduces to

$$(\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}})\, \mathbf{w} = \lambda \boldsymbol{\psi}_{\mathrm{s}} \tag{4.24}$$

The solution for the weights is therefore

$$\mathbf{w} = \lambda\, (\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}})^{-1}\, \boldsymbol{\psi}_{\mathrm{s}} \tag{4.25}$$

The distortionless constraint implies

$$\lambda = \frac{1}{\boldsymbol{\psi}_{\mathrm{s}}^H\, (\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}})^{-1}\, \boldsymbol{\psi}_{\mathrm{s}}} \tag{4.26}$$

which is the MVDR beamforming solution.

As noted in Section 4.2.5, the spatially robust beamformer design also exhibits improved white noise gain characteristics compared with the traditional MVDR beamformer. The robust design has not been explicitly designed for a distortionless response, however the beamforming weights can be normalised such that the response to the mean source angle is undistorted. An interesting comparison between the MVDR and robust beamformers is the response to a source originating from the mean angle of the distribution when sensor noise is present.

In Figure 4.10, the response to the centre of the source distribution is presented for the noisy microphone simulations. Both methods achieve near-distortionless response (less than 1.5 dB difference from the ideal response) throughout the frequency range for the given noise level (0.1%) and matrix regularisation (10^{-6}). The robust method exhibits slightly more consistent response than the MVDR method when the weight vectors are perturbed, as apparent from the confidence intervals in Figure 4.10, a consequence of better white noise gain characteristics.

4.2.8 Conclusion

Modelling a sound source distribution using either the 2D or 3D von Mises(-Fisher) density functions results in simple novel analytic expressions for computing the correlation between microphones, which can be used to design spatially robust beamformers/nullformers capable of tolerating uncertainty in microphone array to source direction. The data in Tables 4.1, 4.2, and 4.3, demonstrate that the broad distribution is capable of tolerating an additional error in the expected direction of arrival subject to an increase in SINR. The spatially robust formulation is particularly well suited for the

application of signal suppression, where the distribution approach allows for a easily specified broad region of suppression — useful for applications where there is uncertainty in the blocking direction. Beamformers based on the distribution approach presented are also more numerically robust than standard beamforming methods, indicating greater tolerance to sensor errors/mismatch.

4.2.9 Proof of 2D von Mises-based Correlation Function

The 2D von Mises density function in (4.9) can be expressed in terms of modified cylindrical Bessel functions, using [Abramowitz and Stegun, 1964, (9.6.34)], as

$$p(\phi, \phi_0) = \frac{1}{2\pi I_0(\kappa)} \left[I_0(\kappa) + 2 \sum_{n=1}^{\infty} I_n(\kappa) \cos(n(\phi - \phi_0)) \right] \tag{4.27}$$

Using Euler's identity and the modified Bessel function property [Abramowitz and Stegun, 1964, (9.6.6)],

$$I_n(x) = I_{-n}(x) \tag{4.28}$$

(4.27) can be expressed as

$$p(\phi, \phi_0) = \frac{1}{2\pi I_0(\kappa)} \sum_{n=-\infty}^{\infty} I_n(\kappa) e^{in(\phi - \phi_0)} \tag{4.29}$$

Inserting (4.29) and (4.8) into (4.2), and using the orthogonality of the complex exponential functions, the correlation function can be expressed as

$$\mathbf{R}_{\mathrm{s}}[a, b] = \frac{1}{I_0(\kappa)} \sum_{n=-\infty}^{\infty} i^n J_n(k r_{ab}) I_n(\kappa) e^{in(\phi_{ab} - \phi_0)} \tag{4.30}$$

which is equivalent to the result in [Teal et al., 2002a].

The modified Bessel functions have the property [Abramowitz and Stegun, 1964, (9.6.3)]

$$I_n(x) = e^{-in\frac{\pi}{2}} J_n(ix) = i^{-n} J_n(ix) \tag{4.31}$$

Noting that

$$i^n = e^{in\frac{\pi}{2}} \tag{4.32}$$

(4.30) can be expressed as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{1}{I_0(\kappa)} \sum_{n=-\infty}^{\infty} J_n(kr_{ab}) J_n(i\kappa) e^{in(\phi_{ab}-\phi_0)} \tag{4.33}$$

Using [Abramowitz and Stegun, 1964, (9.1.79)] and Euler's identity, the summation is reduced to a single term, leading to the simple correlation function

$$\mathbf{R}_{\mathrm{s}}[i,j] = \frac{J_0(z)}{I_0(\kappa)} \tag{4.34}$$

where $z = \sqrt{(kr_{ab})^2 - \kappa^2 - 2i\kappa kr_{ab}\cos(\phi_{ab}-\phi_0)}$.

4.2.10 Proof of 3D von Mises-Fisher-based Correlation Function

Equation (4.14) can be defined in terms of modified spherical Bessel functions using [Abramowitz and Stegun, 1964, (10.2.36)]

$$p(\Omega,\Omega_0) = \frac{1}{4\pi\iota_0(\kappa)} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\Psi_p)\iota_n(\kappa) \tag{4.35}$$

where

$$\cos\Psi_p = \cos\theta\cos\theta_0 + \sin\theta\sin\theta_0\cos(\phi-\phi_0) \tag{4.36}$$

and

$$\iota_n(\kappa) = \sqrt{\frac{\pi}{2\kappa}} I_{n+\frac{1}{2}}(\kappa) \tag{4.37}$$

denotes the n^{th} order modified spherical Bessel function.

Using the spherical harmonic addition theorem [Clapp, 1970, (1.1)] the probability density function in (4.35) can be expressed as

$$p(\Omega, \Omega_0) = \frac{1}{\iota_0(\kappa)} \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \iota_n(\kappa) Y_n^m(\Omega) Y_n^{m*}(\Omega_0) \tag{4.38}$$

Plane-waves in 3D can be described as

$$e^{i\mathbf{k}\cdot\mathbf{r}_a} = 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} i^n j_n(kr_a) Y_n^m(\Omega_a) Y_n^{m*}(\Omega) \tag{4.39}$$

The correlation between two microphones can similarly be described as

$$\begin{aligned} e^{i\mathbf{k}\cdot\mathbf{r}_a} e^{-i\mathbf{k}\cdot\mathbf{r}_b} &= e^{i\mathbf{k}\cdot\mathbf{r}_{ab}} = \psi_a \psi_b^* \\ &= 4\pi \sum_{n=0}^{\infty} \sum_{m=-n}^{n} i^n j_n(kr_{ab}) Y_n^m(\Omega_{ab}) Y_n^{m*}(\Omega) \end{aligned} \tag{4.40}$$

Inserting (4.40) and (4.38) into (4.3),

$$\begin{aligned} \iint_\Omega \frac{4\pi}{\iota_0(\kappa)} \sum_{n_1=0}^{\infty} \sum_{n_2=0}^{\infty} \sum_{m_1=-n_1}^{n_1} \sum_{m_2=-n_2}^{n_2} & i^{n_1} j_{n_1}(kr_{ab}) \iota_{n_2}(\kappa) \\ & Y_{n_1}^{m_1}(\Omega_{ab}) Y_{n_2}^{m_2*}(\Omega_0) Y_{n_1}^{m_1*}(\Omega) Y_{n_2}^{m_2}(\Omega) d\Omega \end{aligned} \tag{4.41}$$

Using the orthogonality property of the spherical harmonics,

$$\iint_\Omega Y_n^{m*}(\Omega) Y_n^m(\Omega) d\Omega = \delta_{n_1 n_2, m_1 m_2} \tag{4.42}$$

(4.41) reduces to

$$\mathbf{R}_{\text{s}}[a, b] = \frac{4\pi}{i_0(\kappa)} \sum_{n=0}^{\infty} \sum_{m=-n}^{\infty} i^n j_n(kr_{ab}) \iota_n(\kappa) Y_n^m(\Omega_{ab}) Y_n^{m*}(\Omega_0) \tag{4.43}$$

which matches the result in [Mammasis and Stewart, 2010].

Using the spherical harmonic addition theorem, the correlation function can be simplified to

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{1}{i_0(\kappa)} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\Psi) i^n j_n(kr_{ab}) \iota_n(\kappa) \tag{4.44}$$

where

$$\cos\Psi = \cos\theta_{ab}\cos\theta_0 + \sin\theta_{ab}\sin\theta_0\cos(\phi_{ab} - \phi_0) \tag{4.45}$$

In a manner similar to the 2D case, it can be shown that the 3D correlation function can be described without a summation as a simple function. The modified spherical Bessel functions can be expressed as

$$\iota_n(x) = e^{-in\frac{\pi}{2}} j_n(ix) \tag{4.46}$$

and so (4.44) can be expressed as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{1}{\iota_0(\kappa)} \sum_{n=0}^{\infty} (2n+1) P_n(\cos\Psi) j_n(kr_{ab}) j_n(i\kappa) \tag{4.47}$$

The summation in (4.47) can be expressed using [Abramowitz and Stegun, 1964, (10.1.45)] as

$$j_0(z) = \sum_{n=0}^{\infty} (2n+1) P_n(\cos\Psi) j_n(kr_{ab}) j_n(i\kappa) \tag{4.48}$$

where $z = \sqrt{(kr_{ab})^2 - \kappa^2 - 2i\kappa kr_{ab}\cos\Psi}$.

The simplified correlation function in 3D is therefore

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{j_0(z)}{\iota_0(\kappa)} \tag{4.49}$$

4.3 Near-field Beamforming

4.3.1 Introduction

This section extends the robust beamforming technique to include near-field distributions of source locations (as opposed to directions). Two distribution models are used to develop the spatial correlation functions: the first of which is a radial Gaussian model; the second of which is a uniform volume distribution.

4.3.2 Source Probability Distribution

The objective is to find a spatial correlation function for the signal received from a distribution of possible source positions centred at the coordinate origin located close to a sensor array rather than assuming a single fixed source position. If it is assumed that the source is located near the coordinate origin then the correlation function for sensors at points r_a and r_b can be obtained by integrating a weighted pair of source to sensor near-field transfer functions over a spherical volume [Grbic et al., 2003].

$$\mathbf{R}_{\mathrm{s}}[a,b] = \int_{\mathrm{vol.}} \zeta_a(r,r_a)\zeta_b^*(r,r_b)p(r,\theta,\phi)dV \tag{4.50}$$

where p denotes the probability distribution function, ζ_a denotes the near-field acoustic pressure at microphone a due to a source at radius r, and the volume element dV is

$$dV = r^2\,dr\,\sin\theta\,d\theta\,d\phi \tag{4.51}$$

Now assume the source distribution is spherically symmetric, i.e., varying only with radius

$$p(r,\theta,\phi) = p_r(r) \tag{4.52}$$

The normalisation conditions require that

$$\int_0^{\infty} \int_0^{\pi} \int_0^{2\pi} p_r(r)\, r^2\, dr\, \sin\theta\, d\theta\, d\phi = 1 \tag{4.53}$$

The Gaussian-like function

$$p_r(r) = \frac{1}{\sqrt{8\pi^3\sigma^3}} \exp\left(-\frac{r^2}{2\sigma^2}\right) \tag{4.54}$$

is a valid solution and meets the imposed criteria to model the probability distribution of the source.

4.3.3 Spatial Correlation Function: Gaussian Distribution

The point source in free space can be expressed in terms of spherical basis functions as [Colton and Kress, 1998] [Williams, 1999]

$$\zeta_a(r, r_a) = -ik \sum_{n=0}^{\infty} \sum_{m=-n}^{n} \gamma_n(kr)\eta_n(kr_a) Y_n^m(\theta, \phi) Y_n^{m*}(\theta_a, \phi_a) \tag{4.55}$$

where γ_n represents the source term, η_n the sensor term, Y_n^m the spherical harmonic functions of order n, m, and the source/sensor elevation and azimuth angles are defined as (θ, ϕ) and (θ_a, ϕ_a) respectively. The source and sensor terms are defined depending on the radii of the source (r) and sensor (r_a) locations from the co-ordinate origin [Colton and Kress, 1998] [Williams, 1999],

$$\gamma_n(kr)\eta_n(kr_a) = \begin{cases} j_n(kr)h_n(kr_a) & \text{if } r < r_a \\ j_n(kr_a)h_n(kr) & \text{if } r \geq r_a \end{cases} \tag{4.56}$$

where j_n denotes the spherical Bessel function of n^{th} order, and h_n denotes the spherical Hankel functions of n^{th} order.

We define the radial component of (4.55), using the conditions in (4.56) as

$$\Lambda_n^1(kr, kr_a) = j_n(kr)h_n(kr_a) \qquad \text{if } r < r_a \tag{4.57}$$

$$\Lambda_n^2(kr, kr_a) = h_n(kr)j_n(kr_a) \qquad \text{if } r \geq r_a \tag{4.58}$$

and the angular component as

$$\Theta_n(\Omega, \Omega_a) = \sum_{m=-n}^{n} Y_n^m(\Omega) Y_n^{m*}(\Omega_a) \tag{4.59}$$

where $\Omega = (\theta, \phi)$.

The point source equation (4.55) can be compactly expressed as

$$\zeta_a(r, r_a, \Omega, \Omega_a) = -ik \sum_{n=0}^{\infty} \Lambda_n(kr, kr_a)\Theta_n(\Omega, \Omega_a) \tag{4.60}$$

with the appropriate superscript of the Λ function depending on the source and sensor radii.

Inserting (4.60) into (4.50) results in an integral with three defined regions corresponding to the cases with sources having radii less than the smallest sensor radius, sources having radii between those of the two sensors, and sources having radii greater than both of the sensors.

$$\begin{aligned}\mathbf{R}_{\mathrm{s}}[a, b] = \sum_{n_a=0}^{\infty} \sum_{n_b=0}^{\infty} \Bigg\{ &\int_0^{r_{\min}} f_{n_a,n_b}^{(1)}\, r^2\, dr + \int_{r_{\min}}^{r_{\max}} f_{n_a,n_b}^{(2)}\, r^2\, dr \\ &+ \int_{r_{\max}}^{\infty} f_{n_a,n_b}^{(3)}\, r^2\, dr \Bigg\} \int_{\Omega} \Theta_{n_a}(\Omega, \Omega_a)\Theta_{n_b}^*(\Omega, \Omega_b)\, d\Omega\end{aligned} \tag{4.61}$$

where the $f_{n_a,n_b}^{(1)}$, $f_{n_a,n_b}^{(2)}$, and $f_{n_a,n_b}^{(3)}$ terms can be derived using the $\Lambda_n^{1,2}$ definitions in (4.57) and (4.58) to give

$$f_{n_a,n_b}^{(1)} = p_r(r)\, h_{n_a}(kr_a)h_{n_b}^*(kr_b)j_{n_a}(kr)j_{n_b}(kr) \tag{4.62}$$

$$f_{n_a,n_b}^{(2)} = p_r(r)\, j_{n_a}(kr_a)h_{n_b}^*(kr_b)h_{n_a}(kr)j_{n_b}(kr) \tag{4.63}$$

$$f^{(3)}_{n_a,n_b} = p_r(r)\, j_{n_a}(kr_a) j_{n_b}(kr_b) h_{n_a}(kr) h^*_{n_b}(kr) \tag{4.64}$$

None of the integrals over r in (4.61) is known to have a simple analytical result. A close approximation is possible if the probability distribution is compact enough such that the correlation function can be represented using only the $f^{(1)}_{n_a,n_b}$ functions. In order to attain an analytical solution the integration limit is extended from $r_{\min}$ to ∞ under the assumption that the additional contribution when integrating from $r_{\min}$ to ∞ is insignificant. That is, we assume that

$$p(r > r_{\min}) \simeq 0 \tag{4.65}$$

which if true implies that

$$\int_0^{r_{\min}} f^{(1)}_{n_a,n_b} r^2\, dr + \int_{r_{\min}}^{\infty} f^{(1)}_{n_a,n_b} r^2\, dr \simeq \int_0^{r_{\min}} f^{(1)}_{n_a,n_b} r^2\, dr \tag{4.66}$$

since it is assumed that the probability of the source location being greater than $r_{\min}$ is approximately zero. In Section 4.3.6 it is demonstrated that this assumption holds for compact distributions of sources near the sensors. The approximate correlation function can therefore be expressed as

$$\begin{aligned} \mathbf{R}_{\mathrm{s}}[a,b] \simeq & \sum_{n_a=0}^{\infty} \sum_{n_b=0}^{\infty} \int_0^{\infty} f^{(1)}_{n_a,n_b}\, r^2\, dr \\ & \times \int_{\Omega} \Theta_{n_a}(\Omega,\Omega_a) \Theta^*_{n_b}(\Omega,\Omega_b)\, d\Omega \end{aligned} \tag{4.67}$$

The angular component

$$\int_{\Omega} \Theta_{n_a}(\Omega,\Omega_a) \Theta^*_{n_b}(\Omega,\Omega_b)\, d\Omega \tag{4.68}$$

can be expanded using (4.59) as

$$\sum_{m_a=-n_a}^{n_a} \sum_{m_b=-n_b}^{n_b} \int_{\Omega} Y^{m_a}_{n_a}(\Omega) Y^{m_b *}_{n_b}(\Omega) Y^{m_a}_{n_a}(\Omega_a) Y^{m_b *}_{n_b}(\Omega_b)\, d\Omega \tag{4.69}$$

The orthogonality of spherical harmonics

$$\int_{\Omega} Y_{n_a}^{m_a}(\Omega) Y_{n_b}^{m_b *}(\Omega)\, d\Omega = \delta_{n_a n_b, m_a m_b} \tag{4.70}$$

can be used to simplify (4.69) to

$$\sum_{m=-n}^{n} Y_n^m(\Omega_a) Y_n^{m*}(\Omega_b) \tag{4.71}$$

The spherical harmonic addition theorem [Abramowitz and Stegun, 1964]

$$\sum_{m=-n}^{n} Y_n^m(\Omega_a) Y_n^{m*}(\Omega_b) = \frac{1}{4\pi}(2n+1) P_n(\cos \Omega_{ab}) \tag{4.72}$$

can be used to further simplify the angular components (where P_n denotes the Legendre polynomials of order n and $\cos \Omega_{ab} = \cos\theta_a \cos\theta_b + \sin\theta_a \sin\theta_b \cos(\phi_a - \phi_b)$ represents the solid angle between the a^{th} and b^{th} sensors in the array).

The approximate correlation function in (4.67) can now be expressed as

$$\mathbf{R}_{\text{s}}[a,b] \simeq \frac{1}{4\pi} \sum_{n=0}^{\infty} (2n+1) P_n(\cos \Omega_{ab}) \int_0^{\infty} f_{n,n}^{(1)}\, r^2\, dr \tag{4.73}$$

Using the definition of the $f_{n,n}^{(1)}$ function in (4.62) and defining

$$\alpha_n = (2n+1) h_n(kr_a) h_n^*(kr_b) P_n(\cos \Omega_{ab}) \tag{4.74}$$

leads to the expression for the approximate correlation function

$$\mathbf{R}_{\text{s}}[a,b] \simeq \frac{k^2}{\sqrt{128\pi^5}\sigma^3} \sum_{n=0}^{\infty} \alpha_n \int_0^{\infty} \exp\left(\frac{-r^2}{2\sigma^2}\right) j_n^2(kr)\, r^2\, dr \tag{4.75}$$

In [Gradshteyn and Ryzhik, 2007, (6.633)] a similar integral solution in terms of cylindrical Bessel functions is given as

$$\int_0^{\infty} \exp(-\beta^2 r^2) J_n(\lambda r) J_n(\mu r)\, r\, dr = \frac{1}{2\beta^2} \exp\left(-\frac{\lambda^2+\mu^2}{4\beta^2}\right) I_n\left(\frac{\lambda\mu}{2\beta^2}\right) \tag{4.76}$$

where I_n denotes the modified cylindrical Bessel functions of order n.

The spherical Bessel functions can be expressed in terms of cylindrical Bessel functions as follows

$$j_n(kr) = \sqrt{\frac{\pi}{2kr}} J_{n+\frac{1}{2}}(kr), \tag{4.77}$$

The integral in (4.75) can therefore be expressed as

$$\frac{\pi}{2k} \int_0^\infty \exp\left(-\frac{r^2}{2\sigma^2}\right) J^2_{n+\frac{1}{2}}(kr) r dr \tag{4.78}$$

which can be solved analytically using (4.76).

Inserting the appropriate constants from (4.78) into (4.76) gives the integral solution

$$\frac{\pi\sigma^2}{2k} \exp\left(-k^2\sigma^2\right) I_{n+\frac{1}{2}}\left(k^2\sigma^2\right) \tag{4.79}$$

Collecting the terms gives the correlation function

$$\mathbf{R}_{\mathrm{s}}[a,b] \simeq \frac{k}{\sqrt{512\pi^3\sigma^2}} \exp\left(-k^2\sigma^2\right) \sum_{n=0}^{\infty} \alpha_n I_{n+\frac{1}{2}}(k^2\sigma^2) \tag{4.80}$$

The modified cylindrical Bessel function can be expressed in terms of the modified spherical Bessel function as follows

$$I_{n+\frac{1}{2}}(k^2\sigma^2) = \sqrt{\frac{2k^2\sigma^2}{\pi}} \iota_n(k^2\sigma^2) \tag{4.81}$$

Substituting into (4.80) gives the simplified correlation function expression

$$\mathbf{R}_{\mathrm{s}}[a,b] \simeq \frac{k^2}{16\pi^2} \exp\left(-k^2\sigma^2\right) \sum_{n=0}^{\infty} \alpha_n \iota_n(k^2\sigma^2) \tag{4.82}$$

Equation (4.82) is presented as an infinite summation over the $\alpha_n\iota_n(k^2\sigma^2)$ terms. In practice, only a few terms are required to compute the correlation function [Li and Duraiswami, 2007], as the higher order Bessel terms require

large values of kr_a to contribute to the sum. For an audio application for example, a compact sensor array ($r_{\text{array}} = 2\,\text{cm}$) for speech (up to 4 kHz) would require just 2 terms to compute an accurate correlation function, using a guideline of $n_{\max} = \lceil kr_{\text{array}} \rceil$ (refer to the spherical Bessel function activation plot Fig. 1b in [Rafaely, 2005], for example).

4.3.4 Spatial Correlation Function: Uniform Distribution

In this section, the near-field source distribution is now assumed to be uniform over some specified volume. The probability distribution in this case is defined as a constant value over some fixed volume defined by a maximum source radius r_s.

$$p_r(r) = \frac{3}{4\pi r_s^3} \tag{4.83}$$

Inserting (4.55) and (4.83) into (4.50), and simplifying through the use of the spherical harmonic addition theorem and orthogonality properties leads to the expression,

$$\mathbf{R}_{\text{s}}[a,b] = \frac{3k^2}{16\pi^2 r_s^3} \sum_{n=0}^{\infty} \alpha_n \int_0^{r_s} j_n^2(kr)\, r^2\, dr, \tag{4.84}$$

where α_n is defined as in (4.74).

The integral

$$\int_0^{r_s} j_n^2(kr)\, r^2\, dr, \tag{4.85}$$

can be evaluated by substituting the definition of the spherical Bessel functions in terms of the cylindrical Bessel functions and evaluating the indefinite integral [Gradshteyn and Ryzhik, 2007, (5.54)] between the integration limits.

$$\frac{\pi}{2k} \int J_{n+\frac{1}{2}}^2(kr)\, r\, dr = \frac{r^2\pi}{4k} \left[J_{n+\frac{1}{2}}^2(kr) - J_{n-\frac{1}{2}}(kr) J_{n+\frac{3}{2}}(kr) \right], \tag{4.86}$$

Re-expressing the integral result in terms of spherical Bessel functions gives

$$\int_0^{r_s} j_n^2(kr)r^2dr = \frac{r_s^3}{2}\left[j_n^2(kr_s) - j_{n-1}(kr_s)j_{n+1}(kr_s)\right] \tag{4.87}$$

provided the source distribution radius does not exceed either of the microphone radii. Inserting this result back into (4.84) results in the simplified correlation function

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{3k^2}{32\pi^2}\sum_{n=0}^{\infty}\alpha_n\left[j_n^2(kr_s) - j_{n-1}(kr_s)j_{n+1}(kr_s)\right] \tag{4.88}$$

Similarly to (4.82), only a few terms of the summation are required for most practical applications.

4.3.5 Infinitesimally Small Distributions

An interesting case to test the derived spatial correlation functions is the result when the source distribution is infinitesimally small at the coordinate origin, i.e., for the Gaussian derived result, σ is set to zero; for the uniform result, the source distribution radius is infinitesimally small.

Setting σ to zero in (4.82) gives the correlation function

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{k^2}{16\pi^2}\sum_{n=0}^{\infty}\alpha_n\iota_n(0) \tag{4.89}$$

The modified spherical Bessel function evaluated at zero is

$$\iota_n(0) = \begin{cases} 1 & \text{if } n = 0 \\ 0 & \text{if } n > 0 \end{cases} \tag{4.90}$$

Inserting this into the correlation function and evaluating $\alpha_0 = h_0(kr_a)h_0(kr_b)^*$, gives the simple solution

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{k^2}{16\pi^2}h_0(kr_a)h_0^*(kr_b), \tag{4.91}$$

which can be recognised as the correlation due to a single point source, located at the coordinate origin, near the two sensors [Williams, 1999, (6.73)].

The uniform distribution result derived in Section 4.3.4 can also be reduced to the single point source result. Starting with (4.88), consider the zeroth order term, as the higher order terms evaluate to zero for infinitesimally small values of kr_s,

$$\mathbf{R}_{\mathrm{s}}^{(0)}[a,b] = \frac{3k^2}{32\pi^2}\alpha_0 \left[j_0^2(kr_s) - j_{-1}(kr_s)j_1(kr_s)\right] \tag{4.92}$$

Note that the $j_{-1}(kr_s)$ term can be expressed as $-y_0(kr_s)$ [Abramowitz and Stegun, 1964, (10.1.12)] and that for small values of kr_s, the spherical Bessel functions can be approximated [Abramowitz and Stegun, 1964, (10.1.4, 10.1.5)] as

$$j_n(kr) = \frac{(kr)^n}{(2n+1)!!}, \tag{4.93}$$

and

$$y_n(kr) = -\frac{(2n-1)!!}{(kr)^{n+1}}, \tag{4.94}$$

where

$$(2n-1)!! = (1 \times 3 \times 5 \times 7 \times 9 \ldots 2n-1) \tag{4.95}$$

denotes the double factorial. The square bracket term in (4.92) can be approximated as

$$\left[j_0^2(kr_s) - y_0(kr_s)j_1(kr_s)\right] = 1 - \frac{1}{kr_s}\frac{kr_s}{3} = \frac{2}{3} \tag{4.96}$$

Substituting in $\alpha_0 = h_0(kr_a)h_0(kr_b)^*$, the zeroth order spatial correlation function can therefore be expressed as

$$\mathbf{R}_{\mathrm{s}}[a,b] = \frac{3k^2}{32\pi^2}\frac{2}{3}h_0(kr_a)h_0^*(kr_b) = \frac{k^2}{16\pi^2}h_0(kr_a)h_0^*(kr_b) \tag{4.97}$$

which again is equal to the expected spatial correlation function due to a single point source.

4.3.6 Simulation Results

The approximate correlation function for a Gaussian source distribution and the exact correlation function for a uniform source distribution were compared with numerically integrated results to establish whether the solutions were sensible; in particular, whether the approximate Gaussian solution in (4.82) was valid for small variance distributions. The correlation functions between two in-line sensors spaced 2 cm apart were computed for a range of source distribution variance (Gaussian) and radius (uniform) values, and source-centre to array-centre distances. The numerically integrated solutions were computed using MATLAB's inbuilt 'integral' function, and we approximated infinity in (4.61) as 100 m for the Gaussian method, due to floating-point inaccuracies which occur in MATLAB's inbuilt Bessel functions for large values of r.

Gaussian Distributed Source Model

In Figure 4.11 the approximate Gaussian-model correlation function is compared to the numerically integrated solution for a sensor spacing of 2 cm and sensors in line with respect to the coordinate origin ($\cos\Omega_{ab} = 1$). The approximate model matches the integrated solution for σ values up to 4 cm, after which the difference between the approximate model and the numerical solution becomes significant. As noted in Section (4.3.3), this was an expected result, because the source distribution begins to have a significant chance of overlapping (and/or exceeding) the sensor array. In Figure 4.12, a simple top-down plot of 10,000 randomised source locations is presented using a standard deviation one fifth ($\sigma = 4$ cm) of the expected source to array-centre

distance (20 cm) — corresponding to the approximate value for which the expression holds with high precision. It can be seen that there are few instances where the randomised source location lies beyond the radius of the sensors in the array, suggesting that the error in the correlation expression should be minimal. As the distribution broadens, the approximation no longer holds and the error in the expression increases.

In Figure 4.13, correlation functions are computed for a range of mean source to array-centre distances. It can be seen that in general, the analytical solution in (4.82) accurately models source distributions with standard deviations up to around one fifth the mean source to array-centre distance.

Uniform Distributed Source Model

Using the same sensor spacing parameters as the Gaussian case (in-line 2 cm spacing, $\cos\Omega_{ab} = 1$), the uniform model shows excellent agreement with the numerical method (Fig. 4.14) — this was expected as this result is an exact solution for the correlation function. Divergence from the numerical integration result occurs when the distribution overlaps the microphone array, as seen towards the right-hand side of Fig. 4.14. This is an expected result as the correlation function in (4.88) is only valid for source distribution radii up to the radius of the smallest source to microphone distance. For the correlation function computed for Figure 4.12, this occurs at 19 cm, which is where the divergence occurs.

Like the Gaussian case, the uniform distribution correlation function was tested for various mean source to array-centre distances to verify that the analytical solution matched the numerically integrated solutions. In Figure (4.15) the analytical result is compared with the numerically integrated result

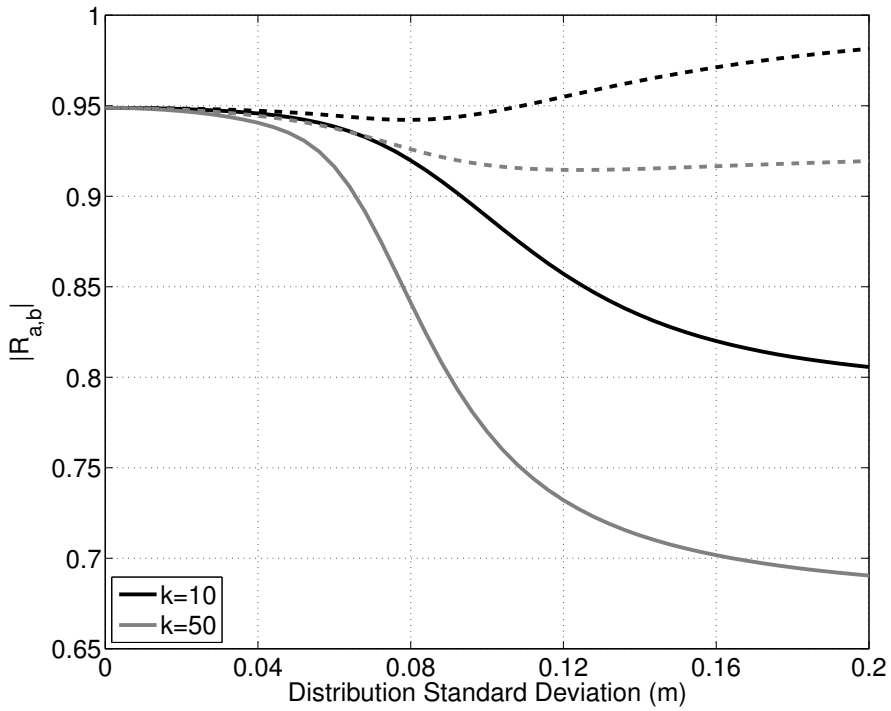

Figure 4.11: Absolute correlation due to Gaussian distributed set of sources computed using (4.82) compared with the numerically integrated result (dashed lines), using a mean source-to-array distance of 20 cm.

for various source-to-array distances. It can be seen that the analytical solution matches the numerical result provided the distribution does not overlap any of the sensors in the array.

4.3.7 Numerical Stability

The solution for the Gaussian distribution presented in (4.82) exhibits a computational problem when calculating the correlation function for high

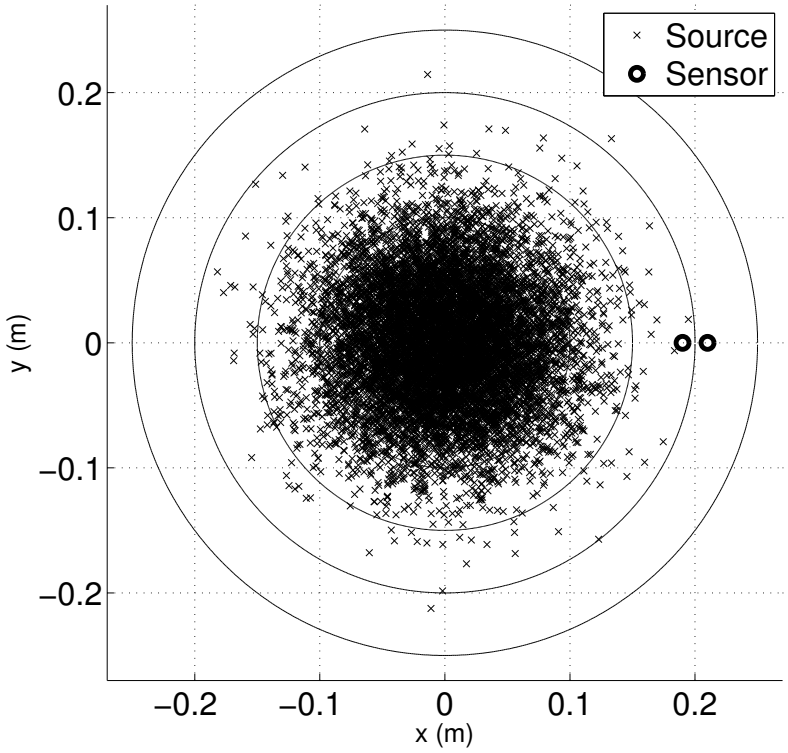

Figure 4.12: An example Gaussian distribution of source locations for $\sigma = 0.04\,\text{cm}$ and a mean source to array-centre distance of 20 cm. A simple 2-sensor line array is pictured to highlight the low probability of source-sensor overlap using a compact distribution.

frequency/wave-number and/or broad distributions. If $k^2\sigma^2$ exceeds ≈ 700, the $\exp(-k^2\sigma^2)$ evaluates to exactly zero in standard 64-bit floating point arithmetic, and the modified spherical Bessel functions, which grow exponentially with increasing $k^2\sigma^2$, evaluate to infinity. It is possible to evaluate the correlation function for large $k^2\sigma^2$ using the series expansion of the modified spherical Bessel functions and Stirling's approximation [Nemes, 2010] for large

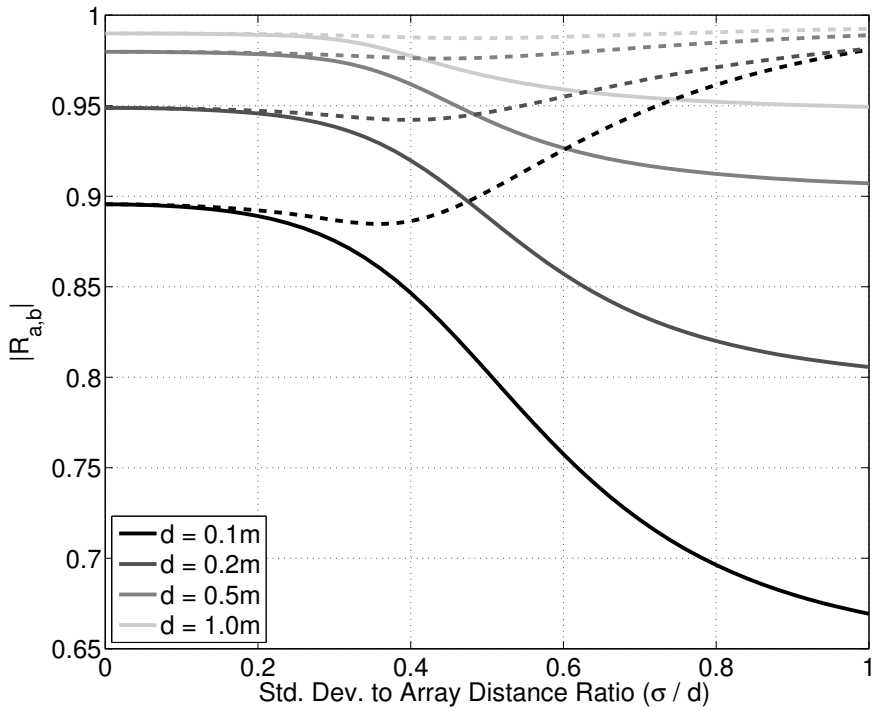

Figure 4.13: Absolute correlation due to a Gaussian distributed set of sources (with wavenumber $k = 10$) using mean source to array-centre distances (d) of 10, 20, 50 and 100 cm. The solid curves represent the analytical function; the dashed curves represent the numerically integrated solution.

factorial/gamma function values.

The series expansion of $\iota_n(x) = \sqrt{\pi/2x}\, I_{n+\frac{1}{2}}(x)$ is given in [Abramowitz and Stegun, 1964, (9.6.10)] as

$$\iota_n(x) = \sqrt{\frac{\pi}{2x}} \sum_{l=0}^{\infty} \frac{1}{l!\Gamma(l + \frac{1}{2} + n + 1)} \left(\frac{x}{2}\right)^{2l+\frac{1}{2}+n} \tag{4.98}$$

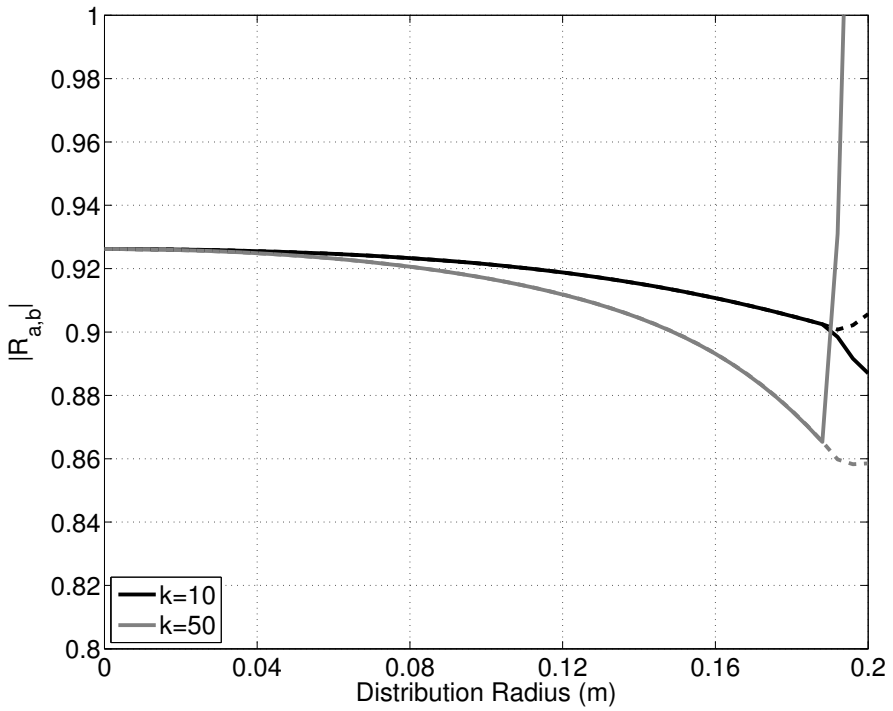

Figure 4.14: Correlation between two sensors due to a uniform distribution of sources. The solid curves represent the solutions given by (4.88), the dashed curves represent the numerically integrated results. Using a source-to-array mean distance of 20 cm.

(4.98) can be expressed as a sum of exponential functions,

$$\iota_n(x) = \sum_{l=0}^{\infty} e^{\beta_l} \tag{4.99}$$

Defining $\mu = l + n + \frac{3}{2}$, $\nu = 2l + n + \frac{1}{2}$, and substituting $k^2\sigma^2$ for x, the

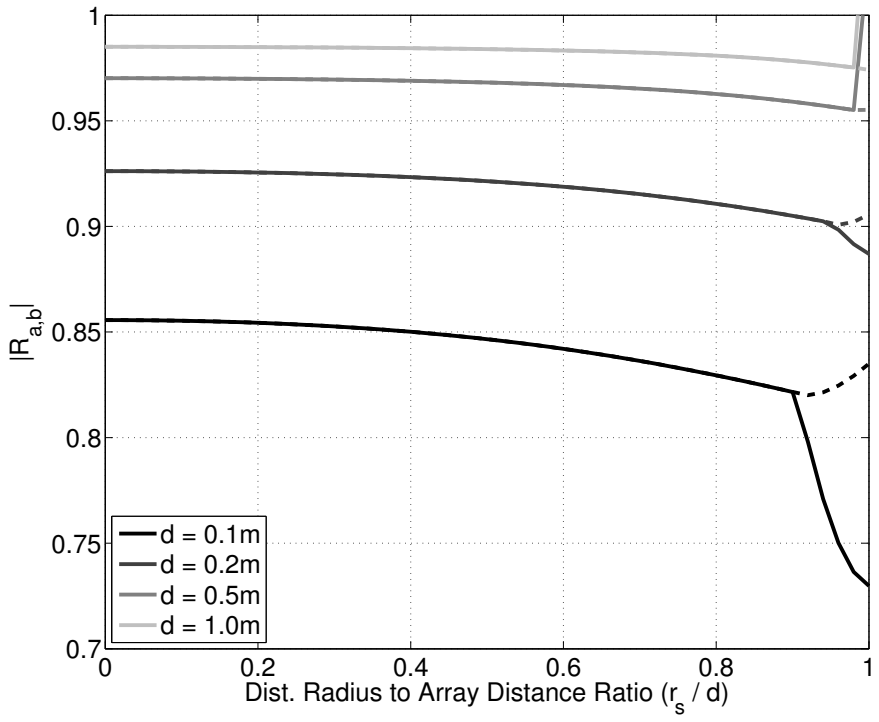

Figure 4.15: Absolute correlation between two sensors as a function of distribution radius due to a uniform distribution of sources (with wavenumber $k = 10$) located 10, 20, 50 and 100 cm from the centre of the sensor array. The solid curves denote the analytical solution, the dashed curves denote the numerically integrated solution.

exponents of each summation term in (4.99) can be computed as follows,

$$\beta_l = \frac{1}{2}\log\left(\frac{\pi}{2k^2\sigma^2}\right) + \nu\left(\log(k^2\sigma^2) - \log(2)\right) - \left(\log(l!) + \log(\Gamma(\mu))\right) \tag{4.100}$$

where the factorial and gamma functions can be computed using Stirling's approximation.

The product of the decaying exponential function and the modified spherical Bessel function can be approximated as

$$e^{-k^2\sigma^2}\iota_n(k^2\sigma^2) = \sum_{l=0}^{\infty} e^{\beta_l - k^2\sigma^2} \tag{4.101}$$

Using this method, it is possible to very rapidly and reliably compute the correlation function for very high frequencies and/or large distribution variances.

4.3.8 Application: Microphone Beamforming

A simple application of the spatial correlation result is the design of a robust microphone beamformer [Martinez et al., 2015] or target suppressing nullformer (used for example, to estimate background noise [Anderson et al., 2015a]). Using the generalised eigenvalue beamformer described in [Anderson et al., 2015a] [Shahbazpanahi et al., 2003], both robust maximum and minimum SINR beamformers can be designed.

The maximum/minimum SINR beamformers can be designed by maximising/minimising the Rayleigh quotient representing the expected output SINR of beamformed signals, described in Section 2.3.6. Repeating the SINR definition in (2.98) here as

$$\text{SINR} = \frac{\mathbf{w}^H \mathbf{R}_\text{s} \mathbf{w}}{\mathbf{w}^H \left[\mathbf{R}_\text{v} + \mathbf{R}_\text{n}\right] \mathbf{w}} \tag{4.102}$$

where $\mathbf{w}$ denotes a vector of narrowband beamformer weights to solve for, $\mathbf{R}_\text{s}$ is the source spatial correlation matrix with entries defined using (4.82) or (4.88); $\mathbf{R}_\text{v}$ the interference correlation matrix — which under a 3D diffuse interference assumption has entries defined earlier in (2.119) as

$$\mathbf{R}_\text{v}[a,b] = j_0(kr_{a,b}) \tag{4.103}$$

where k is the wavenumber and $r_{a,b}$ is the inter-element distance; and $\mathbf{R}_{\mathrm{n}}$ the sensor noise correlation matrix — which is usually defined as

$$\mathbf{R}_{\mathrm{n}} = \sigma_{\mathrm{s}}^2 \mathbf{I} \tag{4.104}$$

The minimum/maximum SINR beamformer solution can be attained by solving (2.102) restated here as

$$\left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{q} = \lambda \mathbf{R}_{\mathrm{s}} \mathbf{q} \tag{4.105}$$

where $(\lambda, \mathbf{q})$ are the solution eigenvalue/vector pairs. The maximium SINR beamformer can be found by selecting the eigenvector associated with the largest eigenvalue; the minimum SINR beamformer can be found by selecting the eigenvector associated with the smallest eigenvalue.

In Figures 4.16 and 4.17, beam/nullformer response patterns for a simple compact 4-element, 2 cm radius circular microphone array with an expected source-to-array centre distance of 20 cm and varying values of distribution variance/width are presented. Using the spatial correlation functions from (4.82) and (4.88) in the generalised eigenvalue beamformer formulation, results in improved spatial robustness, in particular when designing a nullformer (as seen in the x-y plane in Figures 4.18 and 4.19), compared with the cases where the variance/source radius is zero. The result is similar to previous work in robust far-field beamforming in Section 4.2, in which improvement in spatial robustness for the beamformers corresponds to improved suppression of background diffuse interference outside the region of interest. Of note is the asymmetry of the nullformer responses in Figures 4.16 and 4.17 — these result from the poor conditioning of the $\mathbf{R}_{\mathrm{s}}$ matrix in (4.105). The similar patterns for the nullformer responses is a result of the choice of σ and r_{s}

chosen for evaluation. The Gaussian method does tend to produce a slightly deeper and narrower suppression region compared with the uniform method. This is due to the greater assumed source location density near the target position for the Gaussian design.

4.3.9 Application: Simple Adaptive Filtering vs. GSC

As in Section 4.2.6, a simple adaptive filtering scheme was compared to the GSC beamformer to evaluate performance with location mismatch errors. A desired source (speech) was placed 30 cm away from a 4-element 1 cm radius circular array, with an interferer (non-stationary speech/music) placed 1 m away from the array directly opposite the desired source. The mean input SINR was set to 0 dB and sample rate was 44.1 kHz. The adaptive filter lengths were set to 1024 taps, and a step-size parameter of $\mu = 10^{-4}$ was used to control adaptation of the simple filter and GSC designs.

As in the far-field case, the perfect information GSC structure had issues with the blocking matrix design, with the theoretically perfect nullformers in practice not suppressing the target sufficiently well to prevent desired signal distortion effects, which led to a reduction in SINR compared with the robust method — as seen in Table 4.5. However, compared to the far-field example, the GSC structure did tend to always improve SINR over simple fixed beamforming due to near-field gain effects (exploiting relative attenuation between microphones). These effects led to improved nullformer/blocking matrix performance compared with the far-field case, which partly counteracted the numerical accuracy issues with the perfect GSC blocking matrix. Overall, the simple robust method delivered a substantial performance advantage (3-5 dB) over the GSC beamformer due to its improved spatial (and numerical)

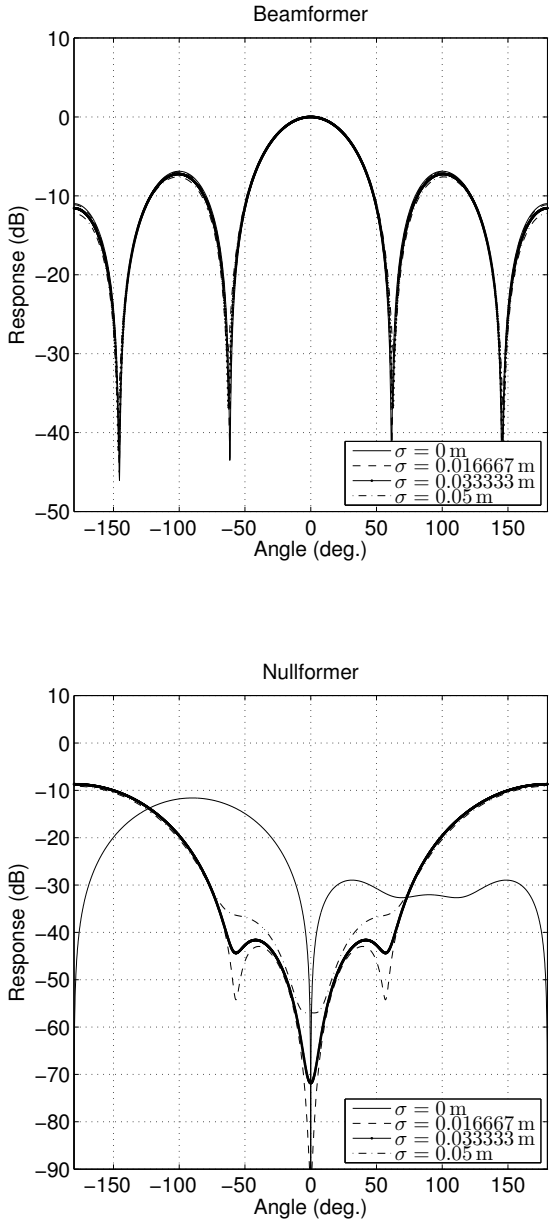

Figure 4.16: Beamformer and nullformer responses in the x-y plane at a distance of 20 cm designed using the Gaussian spatial correlation function with a design frequency of 2.5 kHz.

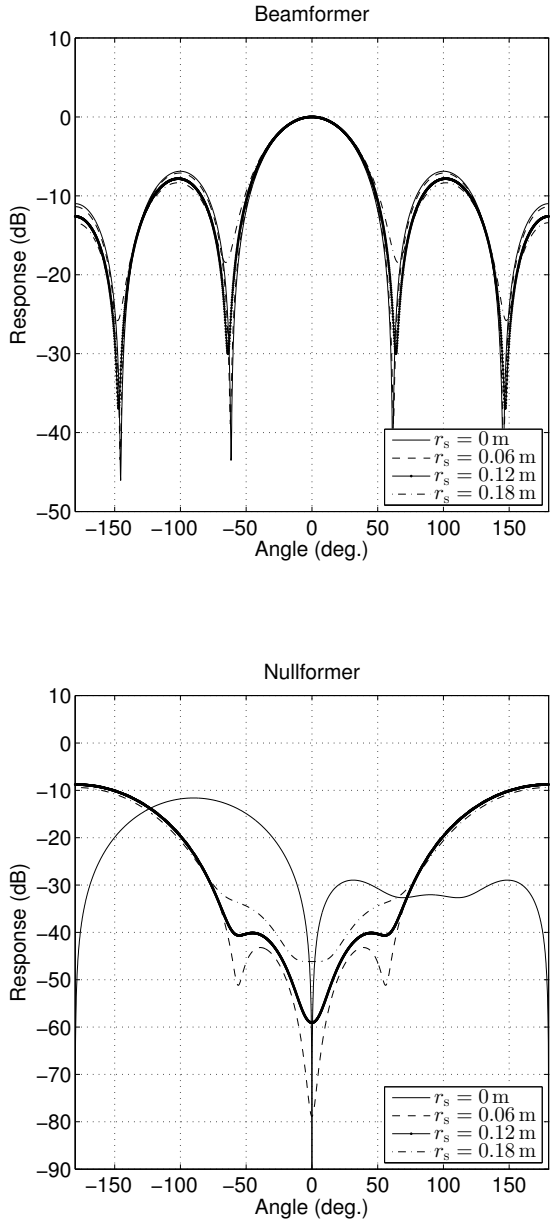

Figure 4.17: Beamformer and nullformer responses in the x-y plane at a distance of 20 cm designed using the uniform spatial correlation function with a design frequency of 2.5 kHz.

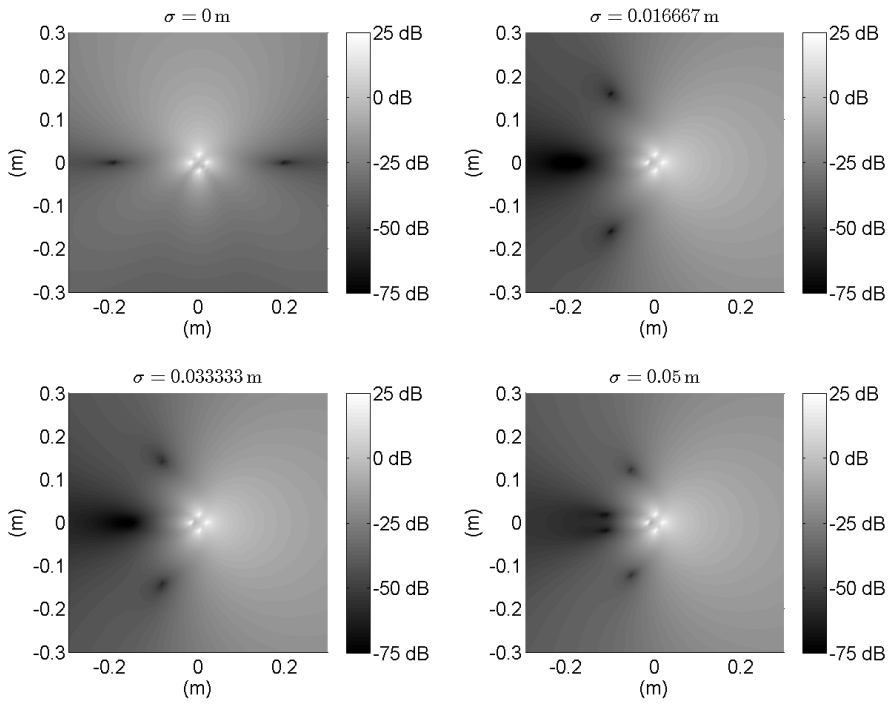

Figure 4.18: Nullformer responses at 2.5 kHz for varying values of σ using the Gaussian distribution correlation method. The target expected location was (-0.2 m, 0 m).

robustness for this simple example.

4.3.10 Discussion

Initially, an assumption was made that the radial Gaussian source position distribution would not likely overlap and exceed the sensor array. This assumption was primarily made to ensure an analytical result was possible, but also to reflect potential applications of the derived correlation function. In beamforming for example, the correlation function would be useful to design

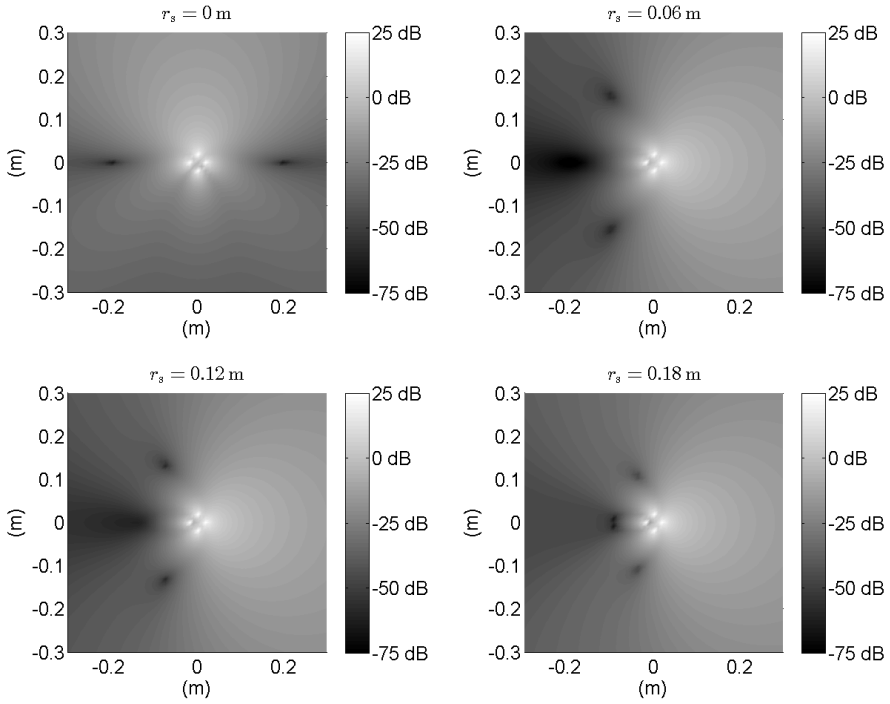

Figure 4.19: Nullformer responses at 2.5 kHz for varying values of r_s using the uniform distribution correlation method. The target expected location was (-0.2 m, 0 m).

a spatially robust beamformer. In such a scenario, the assumption restricts the target source from being inside or behind (opposite the expected location) the sensor array — the first example is unrealistic for many compact array applications (corresponding to the source having some probability of being inside the array); the second corresponds to an increasingly diffuse source scenario, since the source distribution would have a significant probability of surrounding the array, which can be described by the well known sinc function [Teal et al., 2002b].

Table 4.5: SINR comparison between the simple adaptive filter system, GSC, fixed robust, and fixed-MVDR beamformers.

Mismatch	AF (dB)	GSC (dB)	Fixed-Beam (dB)	Fixed-MVDR (dB)
0 cm	14.3	11.1	9.6	9.7
3 cm	14.2	10.9	9.5	9.6
6 cm	13.9	10.1	9.2	9.2
9 cm	13.4	8.8	8.6	8.6
12 cm	12.6	7.4	7.8	7.8
15 cm	11.5	5.8	6.7	6.7

In the simulation results section, we used simple geometry to compare the spatial correlation function solutions and existing numerical methods. The sensor spacing was set to 2 cm and they were placed in-line with the coordinate origin ($\cos\Omega_{a,b} = 1$). For values of $\cos\Omega_{a,b}$ significantly less than 1 (for example, when using larger arrays), we found that the accuracy of the Gaussian solution reduces with respect to numerical methods, compared to the $\cos\Omega_{a,b} = 1$ case, when modelling large source location variances (greater than one-fifth the expected source-to-array distance), this is not a significant issue however, since we have made the compact distribution assumption in order to obtain an analytical result for the spatial correlation function. The uniform spatial correlation solution is exact, as long as the distribution does not overlap the array, and as such does not have any issues.

4.3.11 Conclusion

A pair of spatial correlation functions for spherical distributions of near-field source locations for a Gaussian radial distribution and uniform distribution have been derived without requiring computationally costly numerical integration calculations. It can be seen in one particular application, microphone beamforming, that incorporating these correlation functions into the beamformer design can result in improved spatial robustness, which can be beneficial when there is uncertainty in the source location relative to a sensor array.

Chapter 5

Beamforming with Scatterers

5.1 Outline

This chapter introduces a more realistic acoustic transfer function formulation incorporating scattering and diffraction effects due to a spherical head model.

5.2 Spherical Scatterer Beamforming

5.2.1 Introduction

Scattering by nearby objects is an often neglected issue in beamforming, which may have a significant effect on the performance of beamforming algorithms. A significant number of applications of audio beamforming involve targeting mouths close to a microphone array: phones, notebooks/tablets with integrated webcam/microphone arrays, teleconferencing equipment, and so forth.

Solid sphere scattering/diffraction models have been used in the literature to design microphone beamformers for applications such as headsets [Laugesen et al., 2003] and hearing aids [Merks et al., 2014]. In the latter example, the authors found there was little improvement in the directivity index (equivalent to signal to diffuse interference ratio) when using a scattering model compared with a free-field design. Directional interference, and frequency distortion of the desired signal, introduced by scattering, were not considered in either paper.

In this chapter, the isotropic diffuse interference correlation function is derived for the spherical scatterer model. Additionally, an anisotropic interference correlation function using the von Mises-Fisher distribution is derived. Using these correlation functions, optimal signal to interference plus noise (SINR) beamformers are designed and compared with their free-field equivalents, using SINR and frequency distortion as performance measures. These measures show small improvements over free-field designs.

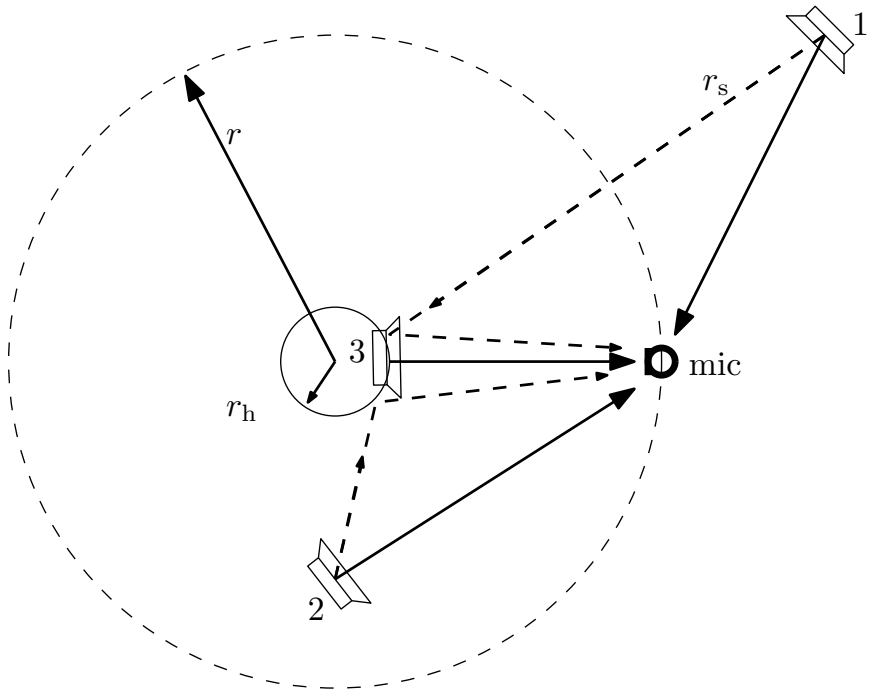

Figure 5.1: Example of near-field sources near a solid sphere. 1) the source outside the microphone radius, 2) the source inside the microphone radius, and 3) a source located on the sphere.

5.2.2 Near-field Source Description

In Figure 5.1 the three scattering scenarios to be considered are presented. The first scenario is the case where a source is located at some radius r_s greater than the sensor radius r in the presence of a solid sphere with radius r_h, which will be used to compute the (non-)isotropic interference correlation functions; the second scenario is the case where the source lies between the

sphere and the sensor; and the third scenario is an extension of the second, in which the source is located on the surface of the sphere.

In general, the equation describing wave propagation from a source to a sensor located in the presence of a solid sphere can be expressed as the combination of a direct path and scattered path.

$$\psi(k, r, r_\mathrm{s}, \Omega, \Omega_\mathrm{s}) = \psi_\mathrm{dir} + \psi_\mathrm{sca} \tag{5.1}$$

The direct components can be expanded in terms of spherical harmonics/spherical Bessel functions as either

$$\psi_\mathrm{dir}(k, r, r_\mathrm{s}, \Omega, \Omega_\mathrm{s}) = \sum_{n,m} A_n^m(k, r_\mathrm{s}, \Omega_\mathrm{s}) j_n(kr) Y_n^m(\Omega) \tag{5.2}$$

for the case where the source radius is greater than the sensor radius (scenario 1 in Figure 5.1) [Williams, 1999, (6.140)], and

$$\psi_\mathrm{dir}(k, r, r_\mathrm{s}, \Omega, \Omega_\mathrm{s}) = \sum_{n,m} A_n^m(k, r_\mathrm{s}, \Omega_\mathrm{s}) h_n(kr) Y_n^m(\Omega) \tag{5.3}$$

for the case where the source radius is less than the sensor radius (scenario 2 in Figure 5.1) [Williams, 1999, (6.92)]. Here, j_n denotes the spherical Bessel function of the first kind, h_n denotes the spherical Hankel function, Y_n^m the spherical harmonics of order n and degree m, $k = 2\pi f/c_0$ the wavenumber (defined using the frequency f and speed of sound c_0), r the sensor radius, r_s the source radius, Ω the sensor angles (θ, ϕ), and Ω_s the source angles (θ_s, ϕ_s). The double sum

$$\sum_{n,m} \equiv \sum_{n=0}^{\infty} \sum_{m=-n}^{n}$$

has been contracted to improve equation clarity.

The scattered components can be expressed as outgoing waves

$$\psi_{\text{sca}}(k, r, r_{\text{s}}, \Omega, \Omega_{\text{s}}) = \sum_{n,m} B_n^m(k, r_{\text{s}}, \Omega_{\text{s}}) h_n(kr) Y_n^m(\Omega) \tag{5.4}$$

For a solid sphere, the scattering coefficients B_n^m can be found by enforcing a zero radial velocity condition on the surface of the sphere [Williams, 1999].

$$\frac{\partial}{\partial r}|_{r=r_{\text{h}}} \left(\psi_{\text{dir}}(k, r, r_{\text{s}}, \Omega, \Omega_{\text{s}}) + \psi_{\text{sca}}(k, r, r_{\text{s}}, \Omega, \Omega_{\text{s}})\right) = 0 \tag{5.5}$$

Relating A_n^m and B_n^m, this condition gives

$$B_n^m(k, r_{\text{s}}, \Omega_{\text{s}}) = -A_n^m(k, r_{\text{s}}, \Omega_{\text{s}}) \frac{j_n^{'}(kr_{\text{h}})}{h_n^{'}(kr_{\text{h}})} \tag{5.6}$$

Point Source Description

The equation describing a near-field point source in free-field (ignoring time-dependence) is

$$\psi_{\text{dir}} = \frac{e^{ik|\mathbf{r}-\mathbf{r}_{\text{s}}|}}{4\pi|\mathbf{r}-\mathbf{r}_{\text{s}}|} \tag{5.7}$$

has the spherical harmonic expansion [Abramowitz and Stegun, 1964, (10.1.45, 10.1.46)]

$$\psi_{\text{dir}} = -ik \sum_{n,m} j_n(kr_-) h_n(kr_+) Y_n^m(\Omega) Y_n^m(\Omega_{\text{s}})^* \tag{5.8}$$

where the r_- and r_+ terms correspond to the smaller and larger radii of r and r_{s}. Equating (5.8) to (5.2) or (5.3) gives the direct path expressions for the scenarios presented in Figure 5.1.

5.2.3 Sources outside microphone radius

Equating (5.8) with (5.2), the direct path coefficients, A_n^m, for sources outside the sensor radius can be expressed as

$$A_n^m(k, r_{\text{s}} \geq r, \Omega_{\text{s}}) = -ikh_n(kr_{\text{s}}) Y_n^m(\Omega_{\text{s}})^* \tag{5.9}$$

The scattering coefficients B_n^m can be found using (5.9) and (5.6) to give

$$B_n^m(k, r_\mathrm{s}, \Omega_\mathrm{s}) = ikh_n(kr_\mathrm{s})\frac{j_n^{'}(kr_\mathrm{h})}{h_n^{'}(kr_\mathrm{h})}Y_n^m(\Omega_\mathrm{s})^* \tag{5.10}$$

The expression for the total field is therefore

$$\psi(k, r, r_\mathrm{s}, \Omega, \Omega_\mathrm{s}) = -ik\sum_{n,m}\left[j_n(kr) - \frac{j_n^{'}(kr_\mathrm{h})}{h_n^{'}(kr_\mathrm{h})}h_n(kr)\right]h_n(kr_\mathrm{s})Y_n^m(\Omega)Y_n^m(\Omega_\mathrm{s})^* \tag{5.11}$$

Sources inside the microphone radius

The A_n^m coefficients for sources inside the sensor radius can be expressed as

$$A_n^m(k, r_\mathrm{s} < r, \Omega_\mathrm{s}) = -ikj_n(kr_\mathrm{s})Y_n^m(\Omega_\mathrm{s})^* \tag{5.12}$$

and the scattering coefficents are given by (5.10).

The expression for the total field is therefore

$$\psi(k, r, r_\mathrm{s}, \Omega, \Omega_\mathrm{s}) = -ik\sum_{n,m}\left[j_n(kr_\mathrm{s}) - \frac{j_n^{'}(kr_\mathrm{h})}{h_n^{'}(kr_\mathrm{h})}h_n(kr_\mathrm{s})\right]h_n(kr)Y_n^m(\Omega)Y_n^m(\Omega_\mathrm{s})^* \tag{5.13}$$

Sources on the sphere

Finally we consider the specific case of (5.13) corresponding to a point source located on the sphere. Setting r_s to r_h in (5.13) gives the expression

$$\psi(k, r, r_\mathrm{h}, \Omega, \Omega_\mathrm{s}) = -ik\sum_{n,m}\left[j_n(kr_\mathrm{h}) - \frac{j_n^{'}(kr_\mathrm{h})}{h_n^{'}(kr_\mathrm{h})}h_n(kr_\mathrm{h})\right]h_n(kr)Y_n^m(\Omega)Y_n^m(\Omega_\mathrm{s})^* \tag{5.14}$$

The term inside the square brackets can be related to the Wronksian of the spherical Bessel functions of first and second kinds [Abramowitz and

Stegun, 1964, (10.1.6)]

$$\begin{aligned} W\{j_n(x), h_n(x)\} &= j_n(x)h_n^{'}(x) - j_n^{'}(x)h_n(x) \\ &= j_n j_n^{'} + ij_n y_n^{'} - j_n^{'} j_n - ij_n^{'} y_n \\ &= iW\{j_n(x), y_n(x)\} \\ &= \frac{i}{x^2} \end{aligned} \tag{5.15}$$

Using the Wronskian identity in (5.15), equation (5.14) can be simplified to

$$\psi(k, r, r_{\mathrm{h}}, \Omega, \Omega_{\mathrm{s}}) = \frac{1}{kr_{\mathrm{h}}^2} \sum_{n,m} \frac{h_n(kr)}{h_n^{'}(kr_{\mathrm{h}})} Y_n^m(\Omega) Y_n^m(\Omega_{\mathrm{s}})^* \tag{5.16}$$

5.2.4 Beamforming

As in the previous chapters, the maximum SINR beamformer method described in Section 2.3.6 can be used to design robust beam and nullformers. As before, the beamformers are found by solving generalised eigenvalue beamformer solution (2.102) restated here as

$$\mathbf{R}_{\mathrm{s}}\mathbf{w} = \lambda \left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right] \mathbf{w} \tag{5.17}$$

and selecting the maximum SINR solution (or minimum SINR solution for nullforming).

The next two sections derive the interference spatial correlation matrices used to design the scattering-based beamformers.

5.2.5 Isotropic Far-field Interference

In highly reverberant environments, the interference received at the microphone array can be modeled as a diffuse isotropic field [McCowan and Bourlard,

2003]. The correlation function for a diffuse field originating at some far distance ($r_\mathrm{s} \gg r$) from the sensor array in the presence of a solid sphere scatterer can be derived using the transfer function equation (5.11) in Section 5.2.2. The correlation function between two sensors can be expressed as

$$\mathbf{R}_\mathrm{v}[a,b] = \int_{\Omega_\mathrm{s}} \psi_a \psi_b^* \, d\Omega_\mathrm{s} \tag{5.18}$$

where $d\Omega_\mathrm{s} = \sin\theta_\mathrm{s} \, d\theta_\mathrm{s} \, d\phi_\mathrm{s}$.

Splitting the components of the transfer function into radial and angular components

$$\psi(r, r_\mathrm{s}) = -ik \sum_{n,m} \rho_n(r, r_\mathrm{s}) Y_n^m(\Omega) Y_n^m(\Omega_\mathrm{s})^* \tag{5.19}$$

Expanding,

$$\begin{aligned} \mathbf{R}_\mathrm{v}[a,b] = k^2 \sum_{n_a,m_a} \sum_{n_b,m_b} & \rho_{n_a}(r_a, r_\mathrm{s}) \rho_{n_b}(r_b, r_\mathrm{s})^* \\ & \times \int_{\Omega_\mathrm{s}} Y_{n_a}^{m_a}(\Omega_a) Y_{n_a}^{m_a}(\Omega_\mathrm{s})^* Y_{n_b}^{m_b}(\Omega_b)^* Y_{n_b}^{m_b}(\Omega_\mathrm{s}) \end{aligned} \tag{5.20}$$

Using the orthogonality of spherical harmonics,

$$\int_{\Omega_\mathrm{s}} Y_{n_a}^{m_a}(\Omega_\mathrm{s})^* Y_{n_b}^{m_b}(\Omega_\mathrm{s}) \, d\Omega_\mathrm{s} = \delta_{n_a n_b, m_a m_b} \tag{5.21}$$

and the spherical harmonic addition theorem [Whittaker and Watson, 1996, (p. 395)],

$$\sum_{m=-n}^{n} Y_n^m(\Omega_a) Y_n^m(\Omega_b)^* = \frac{2n+1}{4\pi} P_n(\cos\Omega_{a,b}) \tag{5.22}$$

where

$$\cos\Omega_{a,b} = \cos(\theta_a)\cos(\theta_b) + \sin(\theta_a)\sin(\theta_b)\cos(\phi_a - \phi_b) \tag{5.23}$$

denotes the angle between sensors a and b, the correlation function (5.20) simplifies to

$$\mathbf{R}_\mathrm{v}[a,b] = \frac{k^2}{4\pi} \sum_{n=0}^{\infty} \rho_n(r_a, r_\mathrm{s}) \rho_n(r_b, r_\mathrm{s})^* (2n+1) P_n(\cos\Omega_{a,b}) \tag{5.24}$$

Assuming the interference is far-field ($r_\mathrm{s} \gg r$), the source-related spherical Hankel functions in (5.11) simplify to

$$h_n(kr_\mathrm{s}) \approx i^n \frac{e^{ikr_\mathrm{s}}}{kr_\mathrm{s}} \tag{5.25}$$

leading to the simplified radial component expression:

$$\rho_n(r, r_\mathrm{s}) = i^{n+1} \frac{e^{ikr_\mathrm{s}}}{kr_\mathrm{s}} [j_n(kr) - \gamma_n(kr_\mathrm{h}) h_n(kr)] \tag{5.26}$$

Inserting (5.26) into (5.24) and defining

$$\gamma_n(kr_\mathrm{h}) \equiv \frac{{j_n}'(kr_\mathrm{h})}{h_n'(kr_\mathrm{h})} \tag{5.27}$$

the correlation function can be expressed as

$$\begin{aligned} \mathbf{R}_\mathrm{v}[a,b] = \frac{1}{4\pi r_\mathrm{s}^2} \sum_{n=0}^{\infty} (2n+1) P_n(\cos \Omega_{a,b}) \\ \times [j_n(kr_a) j_n(kr_b) - \gamma_n(kr_\mathrm{h}) h_n(kr_a) j_n(kr_b) \\ - \gamma_n(kr_\mathrm{h})^* j_n(kr_a) h_n(kr_b)^* + \|\gamma_n(kr_\mathrm{h})\|^2 h_n(kr_a) h_n(kr_b)^*] \end{aligned} \tag{5.28}$$

5.2.6 Directional Far-field Interference

In many scenarios, the interference is directional in nature. The interference correlation function can be modelled by applying a non-uniform probability weighting for each angle

$$\mathbf{R}_\mathrm{v}[a,b] = \int_{\Omega_\mathrm{s}} p(\Omega_0, \Omega_\mathrm{s}) \psi_a \psi_b^* \, d\Omega_\mathrm{s} \tag{5.29}$$

where $p(\Omega_0, \Omega_\mathrm{s})$ is some probability distribution centred at Ω_0.

The von Mises-Fisher distribution [Mammasis and Stewart, 2010] describes a Gaussian-like distribution of sources located on a sphere. The expression of

the probability density function in terms of the distribution spread (κ) and mean direction of arrival (Ω_0) is given by

$$p(\Omega_\mathrm{s}, \Omega_0) = \frac{1}{\iota_0(\kappa)} \sum_{n,m} \iota_n(\kappa) Y_n^m(\Omega_\mathrm{s}) Y_n^m(\Omega_0)^* \tag{5.30}$$

where ι_n denotes the modified spherical Bessel functions of order n.

The interference correlation function can be computed in a manner similar to the method in (5.18), with the addition of the probability density function term.

$$\begin{aligned}\mathbf{R}_\mathrm{v}[a,b] = \frac{1}{\iota_0(\kappa)} \sum_{n_a,m_a} \sum_{n_b,m_b} \sum_{n_p,m_p} \rho_{n_a}(r_a, r_\mathrm{s}) \rho_{n_b}(r_b, r_\mathrm{s})^* \iota_n(\kappa) \times \\ Y_{n_1}^{m_1}(\Omega_a) Y_{n_b}^{m_b}(\Omega_b)^* Y_{n_p}^{m_p}(\Omega_0)^* \int_{\Omega_\mathrm{s}} Y_{n_a}^{m_a}(\Omega_\mathrm{s}) Y_{n_b}^{m_b}(\Omega_\mathrm{s})^* Y_{n_p}^{m_p}(\Omega_\mathrm{s})\, d\Omega_\mathrm{s}\end{aligned} \tag{5.31}$$

The triple spherical harmonic product integral has a solution in terms of Clebsch-Gordan coefficients [Shabtai and Rafaely, 2014].

$$\begin{aligned}\int_{\Omega_\mathrm{s}} Y_{n_a}^{m_a}(\Omega_\mathrm{s}) Y_{n_b}^{m_b}(\Omega_\mathrm{s})^* Y_{n_p}^{m_p}(\Omega_\mathrm{s})\, d\Omega_\mathrm{s} \\ = \sqrt{\frac{(2n_a+1)(2n_p+1)}{4\pi(2n_b+1)}} C_{n_a,n_b,n_p}^{0,0,0} C_{n_a,n_p,n_b}^{m_a,m_p,m_b}\end{aligned} \tag{5.32}$$

The Clebsch-Gordan coefficients have analytic solutions detailed in [Shabtai and Rafaely, 2014, (25)] [Abramowitz and Stegun, 1964, (27.9.1)] as

$$\begin{aligned}C_{n_a,n_b,n_p}^{m_a,m_b,m_p} = \delta_{m_p,m_a+m_b} \sqrt{2n_p+1} \\ \times \left[\frac{(n_p+n_a-n_b)!(n_p-n_a+n_b)!}{(n_a+n_b+n_p+1)!(n_a-m_a)!} \frac{(n_a+n_b-n_p)!(n_p+m_p)!(n_p-m_p)!}{(n_a+m_a)!(n_b-m_b)!(n_b+m_b)!} \right]^{\frac{1}{2}} \\ \times \sum_l \frac{(-1)^{l+n_b+m_b}}{l!} \frac{(n_b+n_p+m_a-l)!}{(n_p-n_a+n_b-l)!} \frac{(n_a-m_a+l)!(n_a-m_a+l)!}{(n_p+m_p-l)!(l+n_a-n_b-m_p)!}\end{aligned} \tag{5.33}$$

for each integer l such that $(n_b + n_p + m_a - l), (n_p - n_a + n_b - l), (n_a - m_a + l), (n_a - m_a + l), (n_p + m_p - l)$, and $(l + n_a - n_b - m_p)$ are all greater than zero.

5.2.7 Results

Isotropic Interference

An important beamformer application is suppressing reverberation, which is commonly modelled as diffuse interference [McCowan and Bourlard, 2003]. Using the beamformer solution in Section 5.2.4 targeting a source located on the sphere and the spatial correlation matrix from Section 5.2.5, the optimal beamformer for suppressing diffuse interference can be obtained.

Using a compact 3D open sphere microphone array consisting of 4-elements with a radius of 1 cm the optimal beamformer (incorporating scattering information) was compared with the standard free-field solution prevalent in the literature. The source to microphone array centre distance was 20 cm and the sphere radius was set to 8.75 cm.

The source correlation matrix $\mathbf{R}_{\text{s}}$ was computed as the outer product of the direct transfer function vector describing the point source to microphone array propagation:

$$\mathbf{R}_{\text{s}} = \boldsymbol{\psi}\boldsymbol{\psi}^H \tag{5.34}$$

The sensor noise correlation/regularisation matrix $\mathbf{R}_{\text{n}}$ described in (2.54) was designed under both relatively noiseless ($\sigma^2 = 10^{-9}$) and noisy ($\sigma^2 = 10^{-2}$) conditions.

The results in Figure 5.2 show that including the scattering information provides no significant (< 0.1 dB) improvement in SINR for the scenario

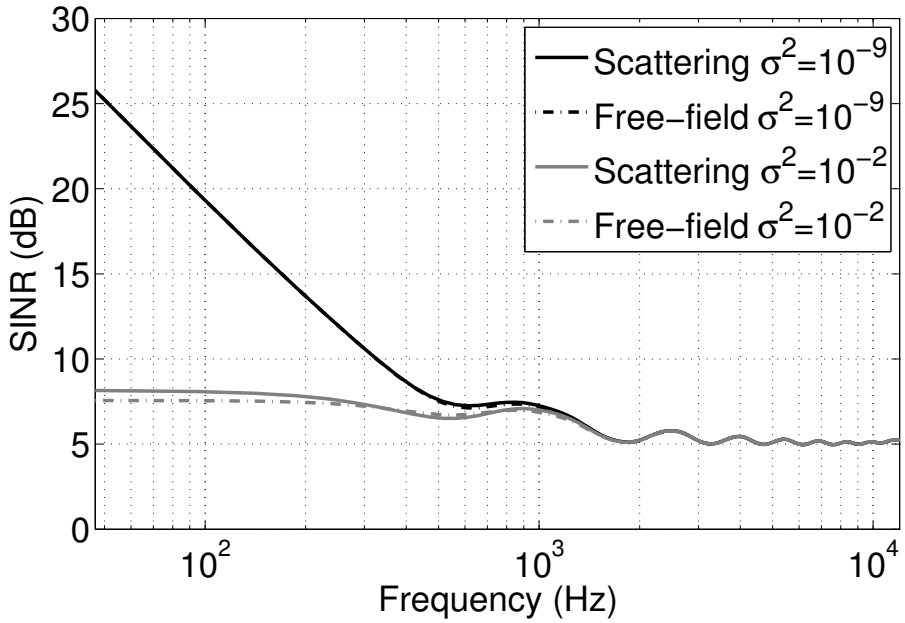

Figure 5.2: SINR Gain (excluding near-field gain) for the scattering information and free-field beamformer designs, for an array in-plane with the source.

of beamforming towards a source located on the sphere (using a source to microphone array distance of 20 cm). This result is similar to the equivalent directivity findings in the literature [Merks et al., 2014].

As identified in Section 2.3.9, white noise gain is a useful measure of beamforming robustness to sensor noise/calibration errors. In Figure 5.3 it can be seen that the WNG improves with increased regularisation. The beamformers designed using a higher sensor noise assumption perform better in terms of WNG as a result.

The beamformers designed using the free-field and scattering formulations were designed to ensure a distortionless response for the desired source under

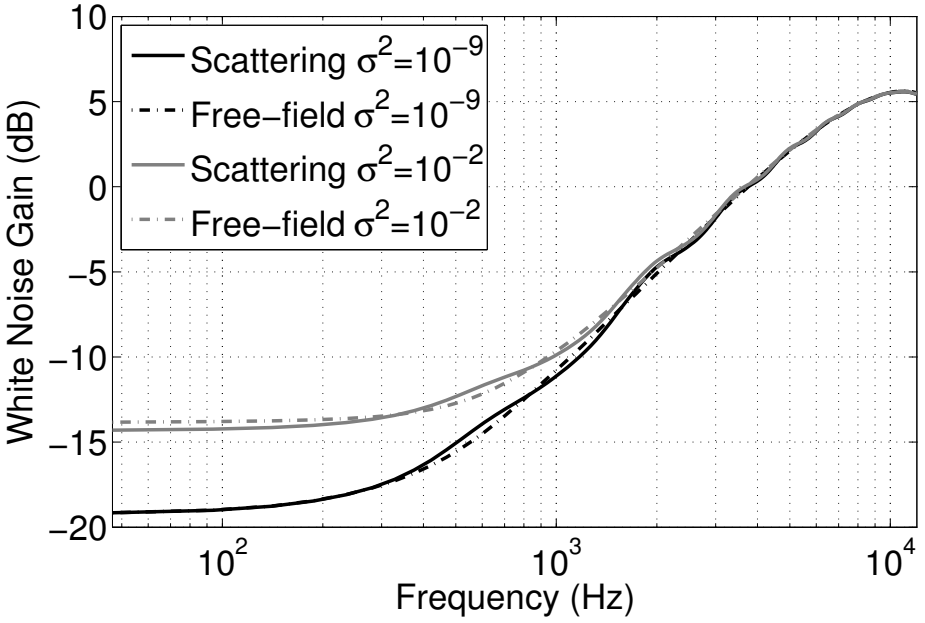

Figure 5.3: White Noise Gain measures for the scatterer-based and free-field beamformer designs, for an array in-plane with the source.

the respective assumed interference conditions. The results in Figure 5.4 demonstrate that using a free-field beamformer design, despite providing the same SINR gain, will introduce a small amount of frequency distortion for the desired source, particularly at low frequencies. This indicates that including the scattering information, although not significantly advantageous in terms of SINR, may be important for applications where a perfectly distortionless signal response is desired. For speech applications however, this distortion would not likely be an issue.

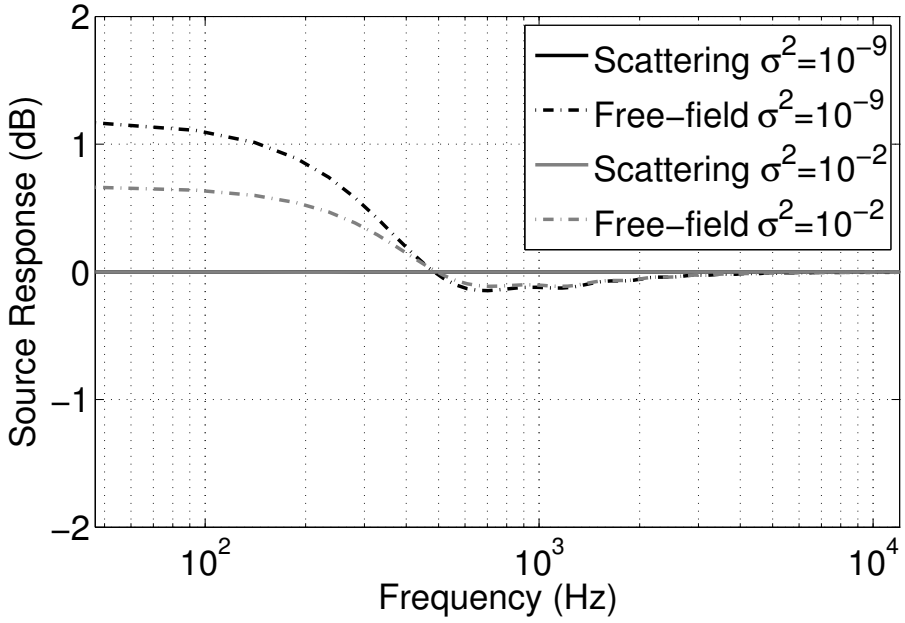

Figure 5.4: Desired signal response for the scattering information and free-field beamformer designs. A distortionless response corresponds to 0 dB.

Directional Interference

For the scenario where interference consists of a directional distribution of sources, the beamformer incorporating scattering information provides significant improvement in SINR, depending on the distribution characteristics (spread and mean direction of arrival). As in the isotropic case, the microphone array was a compact 4-element spherical array with a radius of 1 cm. The interference distribution was centred at $(\theta_0 = 90°, \phi_0 = 60°)$ in plane with the centre of the microphone array and desired source on the sphere.

In Figure 5.5 a comparison of SINR gain is presented for four interferer distributions ranging from near isotropic (the half-power beamwidth $\alpha = 225°$) to near point-like ($\alpha = 6°$). It can be seen that for isotropic-like distributions

($\alpha = 61°, 225°$) there is no significant advantage to including scattering information, as before. For more directional distributions, a significant performance gap between the scatterer-based and free-field designs emerges. The scatterer-based design is capable of an additional 10-20 dB boost in SINR compared with the free-field design for a significant part of the frequency range tested. In Figures 5.6 and 5.7 the far-field response patterns are compared between the two methods. It can be seen that the scatterer-based design produces a deep (> 40 dB attenuation), well-defined null at 60°, which corresponds well to the expected distribution of interferers. The free-field design by contrast produces some attenuation centred at this angle, however it is significantly weaker (20 dB vs. 40 dB) and broader than the scatterer-based design.

5.2.8 Discussion

In the isotropic interference case, the beamforming solution using scattering information provides no significant SINR performance gain over the free-field design if the source is located 180° in-plane relative to the microphone array. This is not an entirely unexpected result as the sphere provides both shadowing effects and competing reflective effects. For example, an interferer located behind the sphere (relative to the microphone array) lies within the shadowing zone of the sphere and is attenuated as a result; while interference originating from in front of the sphere is enhanced by a strong close reflection from the sphere.

For anisotropic interference, the beamformer solution including the scattering information performs significantly better than a free-field design when suppressing a source originating from an approximately known direction. Null

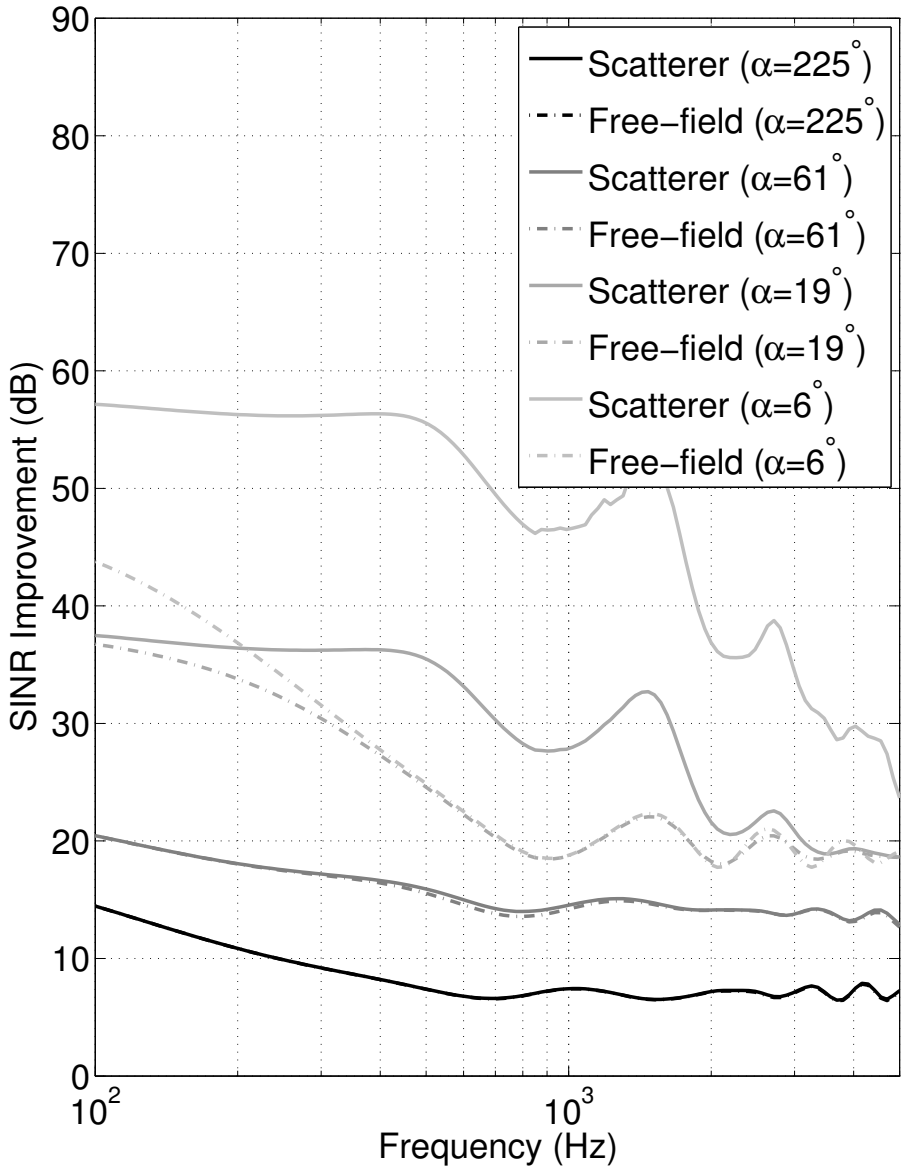

Figure 5.5: SINR improvements for the scattering-based and free-field beamformer designs in the presence of anisotropic interference with a mean direction of 60°.

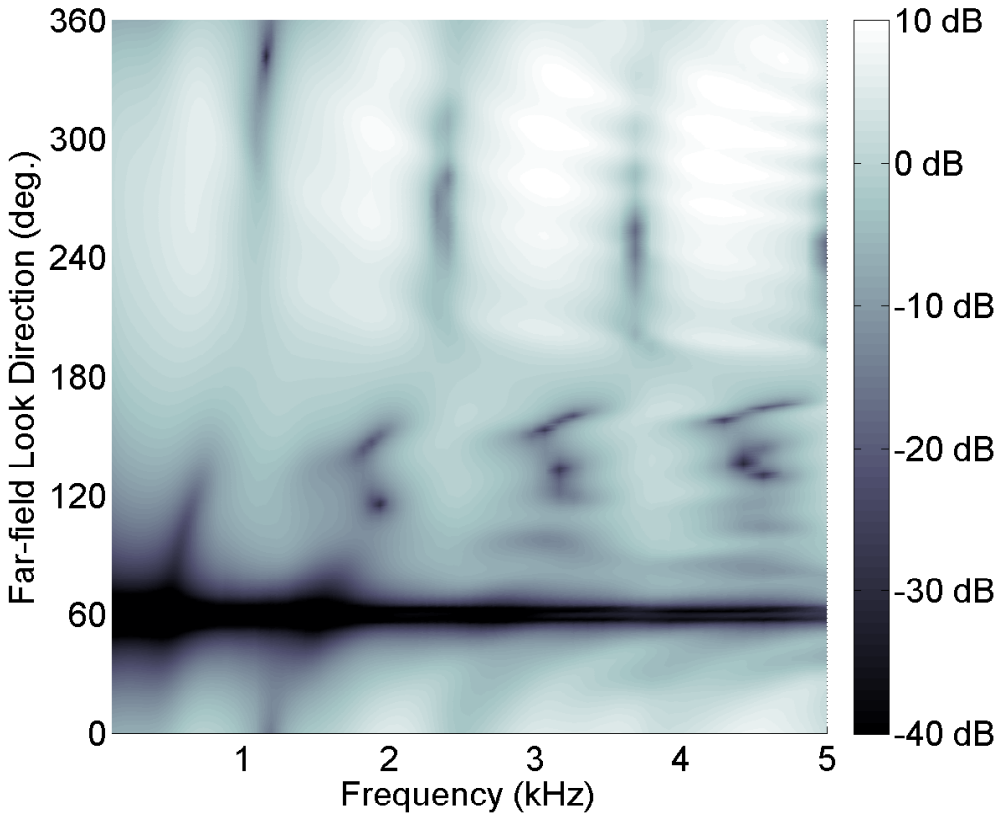

Figure 5.6: Far-field response for the scatterer-based design for a range of frequencies. Interference is centred at 60° with a distribution parameter $\kappa = 500$.

design is sensitive to transfer function mismatch. The spherical scatterer introduces a frequency dependent perturbation to the free-field transfer function. Since the free-field design does not take into account this perturbation, the beamformer solution does not perform as well as the scatterer-based design. This potentially has implications for beamforming algorithms relying

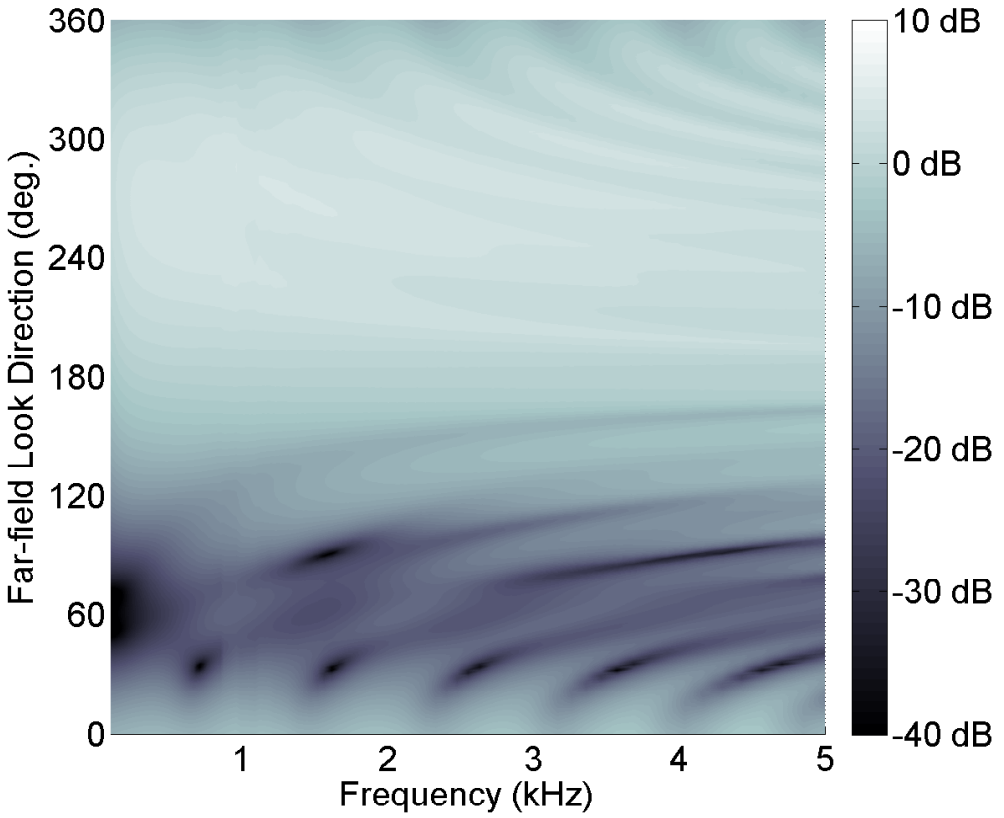

Figure 5.7: Far-field response for the free-field design for a range of frequencies. Interference is centred at 60° with a distribution parameter $\kappa = 500$.

on perfect null creation (such as LCMV and related GSC beamforming) — failure to account (or adapt) for the reflections off the head may lead to degraded performance for these types of algorithms.

5.2.9 Conclusion

The spherical scatterer was introduced as a proxy for modelling the effects of a human head on sound wave propagation. The results for beamforming in the presence of isotropic interference confirm findings of little SINR improvement compared with free-field design. However, the isotropic scenario does benefit from scatterer knowledge to provide distortion compensation to ensure a distortionless response for the desired signal. For the anisotropic interference case, the scatterer information leads to significant improvements in SINR.

It has been suggested that the sphere model is not the most accurate for approximating the effects of a human head [Huopaniemi et al., 1999; Duda et al., 1999], and that a prolate spheroidal model more accurately models scattering effects at higher frequencies (above 2 kHz). However, prolate spheroidal models are considerably more difficult to compute, thus difficult to analyse. As such, we only considered the sphere in this thesis. Future work would include analysis of the prolate model on isotropic and anisotropic interference to see if there are any further advantages over the free-field model. It would be expected that a prolate spheroid beamformer would similarly perform well with anisotropic interference, and may provide a useful base for a robust beamformer design for modelling head related transfer functions (HRTFs) rather than using precise measurements, which in general will differ for each person, and as such may have issues with robustness.

Chapter 6

BSS and Beamforming

6.1 Outline

This chapter describes a complete system utilising dual robust beamformers, which can be designed using any of the techniques from the previous chapters, and a blind source separation post-processor based on the TRINICON algorithm.

6.2 Combined BSS and Beamforming System

6.2.1 Introduction

When applying beamforming for signal extraction, a common objective is to minimise interference while maintaining (ideally) a distortionless response to some desired source. The narrowband signal received at an array of M microphones in the short time frequency domain can be expressed as

$$\mathbf{x}[k,t] = s[k,t]\boldsymbol{\psi}_{\mathrm{s}}[k,t] + \sum_{i=1}^{I} v_i[k,t]\boldsymbol{\psi}_{v,i}[k,t] + \mathbf{n}[k,t], \tag{6.1}$$

where $k = 2\pi f/c_0$ is the wavenumber (where f is the frequency in Hertz, and c_0 is the speed of sound), s and v the desired and interfering signals, $\boldsymbol{\psi}_{\mathrm{s}}$ and $\boldsymbol{\psi}_{\mathrm{v}}$ the $M \times 1$ acoustic transfer function vectors describing the wave propagation from the desired and interfering positions to the microphone locations, and $\mathbf{n}$ represents the sensor noise for each microphone. Ideally, the output of a beamformed system is the undistorted desired signal plus suppressed interference plus noise:

$$y[k,t] = s[k,t] + \mathbf{w}^H[k,t]\left[\sum_{i=1}^{I} v_i[k,t]\boldsymbol{\psi}_{v,i}[k,t] + \mathbf{n}[k,t]\right] \tag{6.2}$$

Assuming the desired signal, interferers and noise are uncorrelated, and of zero mean, the MVDR (Capon) beamformer [Capon, 1969] can be used to generate a beamformer which optimally minimises interference plus noise while maintaining an undistorted response to the desired source location. Dropping the wavenumber and time indexing for clarity, the MVDR solution is given as (as derived in Section 2.3.3)

$$\mathbf{w}_{\mathrm{MVDR}} = \frac{\left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1}\boldsymbol{\psi}_{\mathrm{s}}}{\boldsymbol{\psi}_{\mathrm{s}}^H\left[\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}\right]^{-1}\boldsymbol{\psi}_{\mathrm{s}}}, \tag{6.3}$$

where $\mathbf{R}_{\mathrm{v}} + \mathbf{R}_{\mathrm{n}}$ denotes the interference plus noise spatial correlation matrix.

In most practical scenarios, $\boldsymbol{\psi}_{\mathrm{s}}$ and particularly $\mathbf{R}_{\mathrm{v}}$ are not known precisely and must be estimated to design the beamformer weights. To handle uncertainty in the desired source position, an alternative beamforming solution based on a statistical model of possible desired source locations can be used. The noise spatial correlation matrix is usually estimated by collecting statistics when the desired signal is inactive, which typically involves the use of a voice activity detector for speech applications [Catic et al., 2010]. Noise estimation is usually difficult in low SINR environments, and with multiple non-stationary interferers, so it is sometimes more suitable to use a simpler model of noise spatial correlation to generate the beamformer. In reverberant environments with multiple interferers, an isotropic noise assumption is often appropriate.

More advanced beamforming algorithms attempt to remove residual noise remaining in the output. In Generalised Sidelobe Cancelling (GSC) [Van Trees, 2004] (described in Section 2.3.4), a practical implementation of the MVDR beamformer, a set of orthogonal blocking beamformers, through which the desired signal is suppressed, are used to identify an adaptive filter designed to remove the residual noise. The multichannel Wiener filter, which is equivalent to an MVDR beamformer plus a single-channel Wiener filter post-processor [Van Trees, 2004] (described in Section 2.3.5) is also frequently presented as an optimal method in terms of minimum mean squared error method for noise reduction. Both of these techniques rely, for optimal performance, on precise knowledge of desired signal and/or noise statistics, including the precise location of the desired source. Implementations of these types of algorithms typically rely on training procedures [Gannot et al., 2001; Gannot and Cohen,

2002] to collect the noise correlation statistics. This can be problematic, especially in non-stationary high noise environments [Catic et al., 2010].

In this chapter, an alternative method of noise reduction is presented in which a multiple sensor array is processed via two fixed spatially robust beamformers, a primary beamformer designed to maximise the desired signal to noise ratio, and a second blocking beamformer designed to minimise the desired signal to noise ratio, which are further processed using the TRINICON (Triple-N Independent Component Analysis for Convolutive Mixtures) [Buchner et al., 2004a] blind source separation algorithm as an adaptive processor to correct for inaccurate steering vector and noise statistics assumptions made in the initial design. Previous similar approaches include [Parra and Alvino, 2002], where the authors design a geometrically constrained source separation algorithm, with assumed known precise signal locations. Kumatani et al. [Kumatani et al., 2007] proposed a minimum mutual information-based GSC system for speech separation which avoids the typical signal leakage issues in least squares GSC designs, however their technique also relies on precise target tracking to generate the primary beamformers in their algorithm. This chapter focuses on a spatially fixed simple robust beamforming approach designed to enhance a single desired signal with an uncertain location with uncertain noise correlation statistics. The second-order-statistics version of TRINICON-BSS removes cross-correlations in the output channels, avoiding the target signal cancelling issues inherent in GSC algorithms.

6.2.2 Dual Beamformer Design

The inputs to the TRINICON-BSS system are produced by utilising two beamformers — a primary beamformer which maximises the expected SINR,

and a secondary blocking beamformer which minimises the SINR

$$\text{SINR} = \frac{\mathbf{w}^H \mathbf{R}_\text{s} \mathbf{w}}{\mathbf{w}^H \left[\mathbf{R}_\text{v} + \mathbf{R}_\text{n}\right] \mathbf{w}} \tag{6.4}$$

where $\mathbf{R}_\text{s}$ is the target source spatial correlation matrix, $[\mathbf{R}_\text{v} + \mathbf{R}_\text{n}]$ is the interference plus noise spatial correlation matrix, and $\mathbf{w}$ is the beamforming weight vector to be derived. The beamformer weights can be solved using (2.102) and the appropriate spatial correlation function matrices corresponding to desired signal, interference, and sensor noise.

The optimal beamformer can be designed with the desired source correlation matrix constructed as

$$\mathbf{R}_\text{s,opt} = \sigma_s^2 \boldsymbol{\psi}_\text{s} \boldsymbol{\psi}_\text{s}^H \tag{6.5}$$

and the noise correlation matrix constructed using the expected correlation of the inputs minus the direct desired signal component

$$\mathbf{R}_\text{v+n,opt} = E\left\{\mathbf{x}\mathbf{x}^H\right\} - \mathbf{R}_\text{s,opt} \tag{6.6}$$

which incorporates all interferers, reverberant paths and sensor noise.

Robust beamformers can be generated by utilising probability distribution-based spatial correlation matrices [Dam et al., 2004; Davis et al., 2005]. This formulation assumes that the desired source can be located at any position, with an associated probability distribution. For an arbitrary distribution in spherical coordinates, the spatial correlation matrix entries can be computed using a volume integral

$$\mathbf{R}_\text{s}[a,b] = \int_V p(r,\theta,\phi)\psi_a(r,\theta,\phi)\psi_b^*(r,\theta,\phi)\, dV \tag{6.7}$$

where $p(r,\theta,\phi)$ denotes the source location probability distribution function, and the ψ_a functions denote the wave propagation function from the source to

the a^{th} microphone. For the proposed method, the source location distribution is assumed to be at some fixed distance from the microphone array, sufficient for the far-field source assumption to hold, using a von Mises-Fisher angular distribution to generate the correlation matrix, and assuming free-field anechoic plane wave propagation. From Section 4.2, the desired source spatial correlation matrix can be generated using (4.16). The interference spatial correlation matrix was based on the assumption of isotropically distributed noise sources, including reverberation. Unless specific knowledge of noise distributions in the environment is available, this is a reasonable assumption. For a 3D far-field isotropic case this is given in (2.119) as

$$\mathbf{R}_{\text{v}}[a, b] = j_0(kr_{ab}), \tag{6.8}$$

where j_0 denotes the zeroth order spherical Bessel function.

Typically a Tikhonov regularisation term is included in the noise spatial correlation matrix to improve numerical robustness (corresponding to white noise gain robustness [Cox et al., 1986]), particularly at low frequencies. A regularisation parameter of 10^{-6} was used for the $\mathbf{R}_{\text{n}}$ and $\mathbf{R}_{\text{v+n,opt}}$ matrices.

The primary beamformer ($\mathbf{w}_{\max}$) does not benefit significantly from the robust formulation if the number of microphones and/or array aperture is small — if $kr \leq 1$, where r is the radius of a circular/spherical array for example. From Figure 6.1(a), using the MVDR solution in (6.3) with an assumed mean direction shows almost identical performance to the robust formulation. However, the robust formulation would become quite useful for applications with a large array with a larger number of microphones where the typical MVDR response produces a narrow main lobe.

On the other hand, the use of a distribution of locations is particularly

beneficial in designing the blocking beamformer. In Figure 6.1(b) the expected SINR gain is demonstrated for a perfect null beamformer and a robust nullformer designed using (2.102). It is apparent that sufficient attenuation is only obtained for very small angular regions, whereas the robust method is capable of tolerating a significant uncertainty in the desired source direction. Blocking beamformers used in methods such as the conventional generalised sidelobe canceller (GSC) [Van Trees, 2004] rely on precise nulls, which are not robust to movement. To tolerate perturbations in the desired source direction, GSC implementations require various methods to adapt and track the desired source direction [Gannot et al., 2001; Gannot and Cohen, 2002] which may be unsuitable for high noise environments and/or be computationally expensive. Alternatively, robust GSC implementations such as those presented in [Hoshuyama et al., 1999; Herbordt and Kellermann, 2001] can be used to track the desired source, provided the SINR can be estimated efficiently. The robust nullformer used for the proposed method does introduce some desired signal leakage into the blocking channel, which could lead to filtering issues if they were to be used in GSC-type implementations, which operate by removing correlated components in the blocking path from the primary beamformer channel. The proposed method uses an alternative approach to minimum mean squared error reduction to remove residual noise from the primary beamformer path.

6.2.3 TRINICON-BSS Integration

In Sections 4.2 and 4.3, beamformers were derived using models of signal location and interference correlation. The beamformers were derived using assumptions on the desired signal and interference statistics based on a best

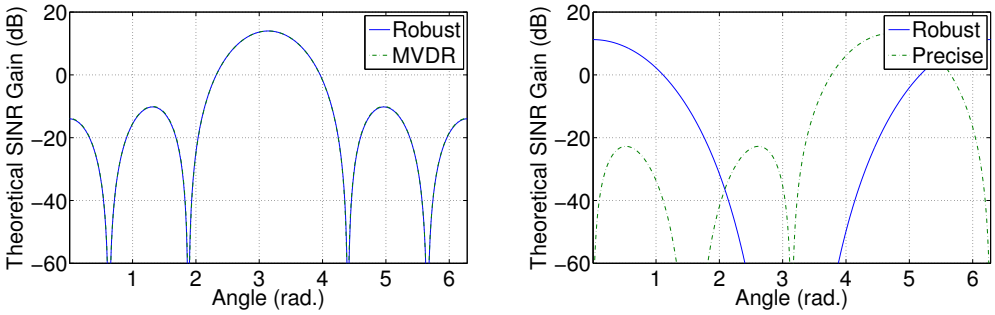

Figure 6.1: Example of a) Expected SINR gain for the robust and perfect MVDR beamformers, and b) Expected SINR gain for the robust and (typical) perfect blocking beamformers

guess of the unknown acoustic scenario. As these assumptions may not accurately represent the scenario, the beamformer performance should be expected to be suboptimal. Integrating a blind source separation algorithm into the system should provide a method for compensating for the assumptions made in the initial beamformer design by exploiting the statistical properties of the beamformer output signals.

The second-order statistics (SOS) version of the TRINICON-BSS algorithm [Aichner et al., 2005] is used to process the beamformer outputs. This BSS algorithm presents many advantages over other frequency-domain BSS algorithms [Hyvärinen, 2001], including the lack of the internal permutation problem — in which the output channel ordering for different frequency bin may not be consistent. The SOS version of TRINICON-BSS also features low computational complexity and can be implemented easily as a real-time algorithm on low-cost, low-power hardware [Aichner et al., 2005; Anderson et al., 2014a].

The cost function for a given block index n in SOS TRINICON-BSS is

given in [Aichner et al., 2005] as

$$J(n) = \sum_{i=0}^{\infty} \beta(i,n) \left[\log\det \mathrm{bdiag}(\mathbf{R}_{yy}(i)) - \log\det \mathbf{R}_{yy}(i) \right], \tag{6.9}$$

where β denotes the block weighting function to incorporate non-stationarity into the algorithm design by including information from the previous blocks (i), $\mathbf{R}_{yy}$ denotes the block-wise output auto/cross-correlation matrix computed from the BSS output channels, and the bdiag operator selects the block diagonal matrices of $\mathbf{R}_{yy}$. This cost function is designed to quantify the level of cross-correlations in the output channels. The gradient-type adaptive filter which minimises this cost function, corresponding to minimising the cross-correlation between the two output channels over all time lags in each block, is specified in [Aichner et al., 2005] as

$$\begin{aligned} \mathbf{W}^{+}_{\mathrm{BSS}}(n) = \mathbf{W}_{\mathrm{BSS}} - \mu \sum_{i=0}^{\infty} \beta(i,n) \\ \times \mathbf{W}_{\mathrm{BSS}} \left[\mathbf{R}_{yy}(i) - \mathrm{bdiag}(\mathbf{R}_{yy}(i)) \right] \mathrm{bdiag}^{-1} \mathbf{R}_{yy}(i), \end{aligned} \tag{6.10}$$

where $\mathbf{W}_{\mathrm{BSS}}$ denotes a Sylvester matrix of filter coefficients, and μ denotes the gradient descent step-size parameter. The Sylvester structure of the filter update and Toeplitz structure of the correlation matrices leads to an efficient frequency domain vector implementation of the algorithm [Aichner et al., 2005; Anderson et al., 2014a]. The implementation used in this chapter uses the block-online design presented in [Aichner et al., 2005], where the β function is approximated by a recursive online function dependent on the parameter λ_{BSS}, set to 0.25, and a block-offline component which iterates the filter update equations five times using the step-size parameter μ set to 0.005. 50% block overlap is used for the BSS algorithm, with the total number of

samples per block (N) set to 3072. The BSS filter length (L) was set to 1024 taps, corresponding to an algorithmic delay of 128 ms when using an 8 kHz sample rate. The regularisation parameters (δ) used in the $\mathbf{R}_{yy}$ block diagonal inverse estimates in (6.10) were set to 10^{-10}.

In the beamformer design, a trade-off was made between desired signal leakage and the angular width for the target suppressing beamformer, introducing desired signal correlation between the two beamformer output channels. The filter updates in the SOS version of TRINICON-BSS (6.10) are designed to remove cross-correlations between the output channels of the overall system, therefore the desired signal leakage should be minimised as part of the separation process.

6.2.4 Simulation Setup

For our experiments, the image source method [Allen and Berkley, 1979] was used to simulate a 6 m × 5 m × 4 m reverberant room with surface reflection coefficients of 0.7, and up to third order reflections used. Four mechanical noise interferers (pump and engine noise) were placed in a circle of radius 3 m centred on the microphone array to simulate isotropic interference. The microphone array was a four-element circular array with 2 cm radius placed in the centre of the room. The desired source, a 30 second sample of speech sampled at 8kHz, was located 1 m from the microphone array. The beamformers were designed for 8kHz wideband signals, with 64-taps for both the robust and optimal beamformers. The implementation of the SOS TRINICON-BSS used in this chapter is identical to that in [Anderson et al., 2014a] using the parameters specified in Section 6.2.3. 50 trials were conducted in which the desired source direction $\phi_s = 180^\circ$ was perturbed by a normally-distributed

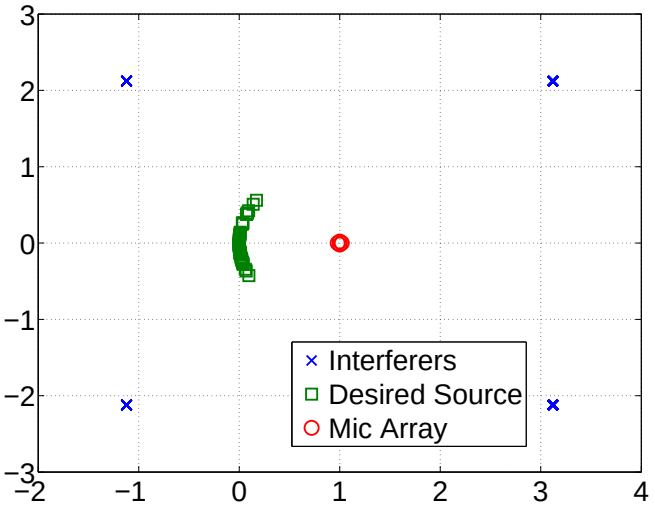

Figure 6.2: Monte Carlo simulation of source positions, interferer locations, and microphone layout

Table 6.1: Mean SINR (dB) results during speech utterances

Peak Input SINR	-6.00	-3.00	0.00	3.00	6.00
Mean Input SINR	-16.94	-13.93	-10.92	-7.91	-4.90
Beamformer	-3.24	-0.24	2.77	5.78	8.79
BF + BSS	-2.07	0.89	3.85	6.81	9.76
Perfect BF	-2.49	0.48	3.44	6.39	9.34

random angle with a standard deviation of $\sigma = 14°$ (Figure 6.2). A further simulation to test channel ordering robustness was conducted in which the desired source position was located at a known fixed location, and the four noise sources allowed to vary position randomly within the room. As in the first case, 50 trials were conducted using the same TRINICON-BSS algorithm parameters in the previous section.

Table 6.2: Mean signal distortion (dB) measures during speech

Peak Input SINR	-6.00	-3.00	0.00	3.00	6.00
Mean Input SINR	-16.94	-13.93	-10.92	-7.91	-4.90
BF + BSS SDR	-25.00	-25.22	-25.49	-25.79	-26.11

6.2.5 Results

The robust beamformer typically results in an improvement of at least 13 dB in terms of SINR for the simulated examples of a peak input SINR of between -6 to 6 dB speech in diffuse noise, as seen in Table 6.1. The inclusion of the blind source separation step improves the mean SINR by up to an additional 3-5 dB during certain speech utterances in the simulations, and on average by 1-1.5 dB over all speech utterances, indicating that this method is able to improve the performance of the array by compensating for some of the assumptions made in the initial beamformer design.

Compared with the perfectly designed (perfect interferer and desired source knowledge) MVDR beamformer, the pre-processed TRINICON system is able to match and sometimes exceed the performance in terms of the SINR gain. The slight performance disadvantage the perfect beamformer exhibits can be attributed to the regularisation introduced into the noise spatial correlation matrix, required for numerical stability, which slightly degrades performance.

The magnitude-squared coherence (MSC)

$$\mathrm{MSC}(k) = \frac{|P_{x_1,y_1}(k)|^2}{P_{x_1,x_1}(k)P_{y_1,y_1}(k)} \tag{6.11}$$

where P_{x_1,y_1} is the cross power spectra density, P_{x_1,x_1} is the input (beamformer channel) power spectra density, and P_{y_1,y_1} is the output (expected speech channel) power spectra density; is a useful measure of separation performance

for BSS algorithms [Fancourt and Parra, 2001]. An MSC value of 1 indicates perfect coherence, while a value approaching zero indicates that the input and output are orthogonal. The modified measure used to evaluate separation performance in this chapter was to take an average of the MSC values for each frequency bin, using a FFT-block size of 3072 samples. The MSC measures in Table 6.3 showing the coherence between the robust beamformer outputs, and the BSS outputs, indicate that there is a reduction in coherence after processing the beamformer outputs through the BSS algorithm. This is an indicator that the BSS algorithm is separating the mixtures. Combined with the SINR results, this suggests that the algorithm is reducing noise in the output channel containing the target signal.

The SINR figures in Table 6.1 show only a small improvement over the robust beamformer, which can be attributed to the negligible improvement in mid to high frequency bins. The robust beamformer is effective at improving the SINR for high frequencies, but performs poorly at low frequencies due to the limited aperture and number of microphones. BSS is able to identify filters which produce a super-directive beamforming effect at low frequencies, however a side-effect of this is that these can be sensitive to microphone position and/or response errors.

The signal distortion measures (the normalised difference in desired signal spectra between the input and output of the system) show that the combined beamforming and BSS algorithm exhibits relatively low desired signal distortion as seen in Table 6.2, with a typical mean value of -25 dB during speech utterances. The BSS process introduces signal distortion, from the on-average undistorted beamformer inputs, into the system by mixing the two beamformer outputs using the BSS filters. There is a small trend towards less

Table 6.3: Integrated MSC measures between the beamformer outputs, and BSS outputs

Peak Input SINR	-6.00	-3.00	0.00	3.00	6.00
Mean Input SINR	-16.94	-13.93	-10.92	-7.91	-4.90
BF Outputs	0.438	0.414	0.390	0.373	0.365
BF + BSS Outputs	0.312	0.281	0.253	0.234	0.225

distortion as the input SINR increases, which is expected as the BSS filters perform less work to decorrelate the outputs. This is also reflected in the SINR results in Table 6.1, where the SINR improvement decreases slightly with increasing input SINR.

In the second simulation designed to test channel ordering robustness, the beamformer plus BSS design exhibited no ambiguity in the output channel ordering. The desired signal was detected consistently in the same first output channel, for the 50 trials. This was an expected result from including the beamformer stage in the system.

6.2.6 Conclusions

A spatially robust adaptive noise reduction algorithm based on spatially robust beamforming and the second-order-statistics version of the TRINICON-BSS algorithm has been presented and compares favourably with a perfect knowledge MVDR beamformer while tolerating significant errors in the assumed desired signal location. By processing the outputs of a robust beam/nullformer pair through BSS, it is possible to compensate for assumptions made in the fixed beamformer design. The algorithm features low signal distortion, fast

convergence and did not exhibit channel ordering ambiguities common in BSS-type algorithms. In addition, the algorithm avoids signal leakage issues common with GSC-type algorithms while maintaining low computational complexity, and does not require speech activity detection, SINR estimation or interference source direction information unlike the existing methods in the literature.

As the method proposed is robust to channel ordering issues, a Wiener filter based post-processor designed using the outputs of the BSS-system, as described in the work in [Reindl et al., 2013], can be easily used to remove residual diffuse noise in the system, leading to a semi-blind multichannel Wiener filter implementation.

Chapter 7

Real-time Implementations

7.1 Outline

This chapter covers real-time implementations of the algorithms developed in the previous chapter. First, a scalable high performance blind source separation system for multiple microphone pairs in described. Next, a complete implementation of the beamforming plus BSS system described in the previous chapter is developed, using a version of the high performance BSS code developed in the first section.

7.2 GPU-Accelerated Blind Source Separation

7.2.1 Introduction

In recent years, graphical processing units (GPUs) have transformed from devices which focus purely on specific tasks relating to 3D graphics processing to general purpose mathematical processing. The advantage that GPUs present over regular CPUs is the ability to process a large amount of data in parallel. A typical GPU provides access to hundreds of processing threads, compared with 2 or 4 in a typical CPU. Signal processing algorithms often involve large filtering operations which are ideal for execution on a massively parallel device. Existing signal processing algorithms which have been modified to run using GPUs include adaptive filtering [Schneider et al., 2012] and Independent Component Analysis (ICA) [Mazur and Mertins, 2011; Foshati and Khunjush, 2013], each of which has demonstrated that significant gains in processing speed are achieved using massively parallel processing. Prior work using GPU acceleration in the field of blind source separation/independent component analysis have focused on algorithms that use frequency bin-wise separation. These algorithms exhibit scaling and permutation problems, limiting their use in audio processing unless repair mechanisms are implemented [Sawada et al., 2004] — which can impair the parallelisation of the algorithms. In addition, little work exists showing the impact of parallelisation and in particular GPU-based algorithms on the accuracy of filter calculations and the resulting effects on separation performance.

One potential application of GPU processing is implementing a quasi-distributed blind source separation system where pairs of microphones (nodes) communicate their audio signals to a centralised computer for processing.

Using a GPU on a central computer may be advantageous if the overheads associated with distributed nodes — efficiencies gained through parallelism, power consumption or CPU time per node compared with the centralised approach, for example, were significant.

7.2.2 Two-Channel BSS Based on TRINICON

TRINICON is a framework for separating convolutive mixtures of signals by exploiting three (assumed) signal properties: non-stationarity, non-whiteness and non-Gaussianity [Buchner et al., 2004b; Aichner et al., 2005]. The objective is to find a set of filters ($\mathbf{W}$) which minimises the cost function imposed by the three signal properties.

Using the formulations given in [Buchner et al., 2004b; Aichner et al., 2005], a two-input/two-output separation example can be given as

$$\begin{bmatrix}\mathbf{Y}_1 & \mathbf{Y}_2\end{bmatrix} = \begin{bmatrix}\mathbf{X}_1 & \mathbf{X}_2\end{bmatrix} \begin{bmatrix}\mathbf{W}_{11} & \mathbf{W}_{12} \\ \mathbf{W}_{21} & \mathbf{W}_{22}\end{bmatrix}, \tag{7.1}$$

where the $\mathbf{Y}_i$ terms are the output channel matrices, $\mathbf{X}_i$ the input channel matrices and the $\mathbf{W}_{ij}$ terms are the separation filter matrices. Each of $\mathbf{X}_i$, $\mathbf{Y}_i$ and $\mathbf{W}_{ij}$ are Toeplitz matrices representing time domain samples and filter taps. The Toeplitz structure of the matrices allows for an efficient vectorised implementation, as detailed in [Aichner et al., 2005].

If the non-Gaussian property assumption is dropped, a second-order statistics-based filter update rule can be utilised to compute the separation filters efficiently. The second-order statistics-based cost function is given in [Aichner et al., 2005] as

$$\mathbf{W}_+ = \mathbf{W} - \mu\mathbf{W}\left[\text{offDiag}(\mathbf{R}_{yy})\text{blockDiag}^{-1}(\mathbf{R}_{yy})\right], \tag{7.2}$$

where μ denotes the gradient descent control parameter, blockDiag is an operator selecting the block diagonal submatrices, offDiag is an operator selecting the off diagonal submatrices and the output correlation matrix $\mathbf{R}_{yy}$ is defined as

$$\mathbf{R}_{yy} = \begin{bmatrix} \mathbf{R}_{yy_{11}} & \mathbf{R}_{yy_{12}} \\ \mathbf{R}_{yy_{21}} & \mathbf{R}_{yy_{22}} \end{bmatrix} = \mathbf{Y}^H\mathbf{Y} = \mathbf{W}^H\mathbf{R}_{xx}\mathbf{W}, \tag{7.3}$$

where $\mathbf{R}_{xx}$ is the input correlation matrix computed from the outer product of $\begin{bmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{bmatrix}^T$. The inverses of the Toeplitz diagonal submatrices of $\mathbf{R}_{yy}$ can be approximated as [Aichner et al., 2006]

$$\mathbf{R}^{-1}_{yy_{ii}} \simeq \frac{1}{\sigma_i^2 + \delta_i}\mathbf{I}, \tag{7.4}$$

where σ_i^2 is the variance of the i^{th} output channel and δ_i is a regularisation parameter to prevent inversion errors.

In [Aichner et al., 2005], the authors implemented an efficient second-order statistics version of this algorithm by exploiting the redundant information in the matrices involved in the equations. The L-length FIR filters in the matrix $\mathbf{W}$ exhibit a Sylvester structure (Figure 7.1) where the columns of each submatrix are diagonally shifted versions of the first column (i.e., time delayed versions of the first column). The submatrices in $\mathbf{R}_{yy}$ also exhibit Toeplitz structure and an approximate inverse can be used for the inverses of the block diagonal terms $\mathbf{R}_{yy_{ii}}$. Their implementation reduces the complexity substantially, allowing real-time separation on low-cost hardware, as the large matrix operations have been replaced by significantly simpler and more memory efficient vector operations.

The fast algorithm is detailed in Table 1 of [Aichner et al., 2005] and summarised as follows, the input time-domain samples are collected as over-

lapping length-N vectors for each channel, windowed and transformed into a frequency-domain vector using the FFT

$$\mathbf{x}_i = \text{FFT}\{\begin{bmatrix} x_i[t] & x_i[t+1] & ... & x_i[t+N-1] \end{bmatrix}^T\} \tag{7.5}$$

The input autocorrelation vectors (representing the first column of each submatrix in $\mathbf{R}_{xx}$) are calculated as element-wise multiplications of the transformed input channels

$$\mathbf{r}_{xx_{ij}} = \mathbf{x}_i \circ \mathbf{x}_j^*, \tag{7.6}$$

where the $\circ$ symbol denotes element-wise multiplication (Hadamard product).

The offline update iteration component can be summarised as the following set of operations: first the current length-L filters are zero-padded to length N and transformed into the frequency domain vector

$$\mathbf{w}_{ij} = \text{FFT}\{\begin{bmatrix} w_{ij}[0] & ... & w_{ij}[L-1] & 0 & ... & 0 \end{bmatrix}^T\} \tag{7.7}$$

In this implementation, the diagonal filters $\mathbf{w}_{ii}$ are initialised to the unit impulse $w_{ii} = \begin{bmatrix} 1 & 0 & ... & 0 \end{bmatrix}$. The off diagonal filters are initialised to zero. The algorithm below operates by altering the off diagonal filters while retaining the diagonal filters as unit impulses — a consequence of the unit impulse initialisation.

The input-output cross-correlations (representing the intermediate step $\mathbf{R}_{xy} = \mathbf{R}_{xx}\mathbf{W}$) can be computed from the input autocorrelation and the current filters and then be used to compute the output autocorrelations

$$\mathbf{r}_{xy} = \begin{bmatrix} \mathbf{r}_{xx_{11}} + \mathbf{w}_{12} \circ \mathbf{r}_{xx_{21}} & \mathbf{r}_{xx_{12}} + \mathbf{w}_{12} \circ \mathbf{r}_{xx_{22}} \\ \mathbf{r}_{xx_{21}} + \mathbf{w}_{21} \circ \mathbf{r}_{xx_{11}} & \mathbf{r}_{xx_{22}} + \mathbf{w}_{21} \circ \mathbf{r}_{xx_{12}} \end{bmatrix} \tag{7.8}$$

$$\mathbf{r}_{yy} = \begin{bmatrix} \mathbf{r}_{xy_{11}} + \mathbf{w}_{12}^{*} \circ \mathbf{r}_{xy_{21}} & \mathbf{r}_{xy_{12}} + \mathbf{w}_{12}^{*} \circ \mathbf{r}_{xy_{22}} \\ \mathbf{r}_{xy_{21}} + \mathbf{w}_{21}^{*} \circ \mathbf{r}_{xy_{11}} & \mathbf{r}_{xy_{22}} + \mathbf{w}_{21}^{*} \circ \mathbf{r}_{xy_{12}} \end{bmatrix} \tag{7.9}$$

The filter updates can be computed efficiently as

$$\mathbf{w} = \mathbf{w} - \mu \begin{bmatrix} \mathbf{0} & \mathbf{r}_{yy_{22}}^{-1} \circ \mathbf{r}_{yy_{12}} \\ \mathbf{r}_{yy_{11}}^{-1} \circ \mathbf{r}_{yy_{21}} & \mathbf{0} \end{bmatrix}, \tag{7.10}$$

where the regularised approximate inverses [Aichner et al., 2006] of $\mathbf{r}_{yy_{ii}}$ are computed by element-wise divisions

$$\mathbf{r}_{yy_{ii}}^{-1} = \frac{1}{\rho \mathbf{r}_{yy_{ii}} + (1-\rho)\sigma_i^2 + \delta_i}, \tag{7.11}$$

with ρ set as a weighting factor and $\sigma_i = \frac{1}{N}\mathbf{r}_{yy_{ii}}^{H}\mathbf{r}_{yy_{ii}}$. The online filter update is computed as

$$\mathbf{w}_{\text{online}} = (1-\lambda)\mathbf{w}_{\text{online}} + \lambda \mathbf{w}, \tag{7.12}$$

where λ is the exponential forgetting factor.

The output channels are computed by convolving the demixing filters with the input channels using the overlap-add FFT convolution method

$$\mathbf{y} = \begin{bmatrix} \mathbf{x}_1 + \mathbf{w}_{12} \circ \mathbf{x}_2 \\ \mathbf{x}_2 + \mathbf{w}_{21} \circ \mathbf{x}_1 \end{bmatrix} \tag{7.13}$$

7.2.3 CUDA

CUDA (Compute Unified Device Architecture) [NVIDIA, 2013], a programming framework for developing GPU-accelerated software for computers with NVIDIA graphics hardware, was chosen for the GPU implementation. The CUDA framework includes a parallel implementation of the FFT similar in style to the *fftw* [FFTW, 2013] library available for CPUs. All of the

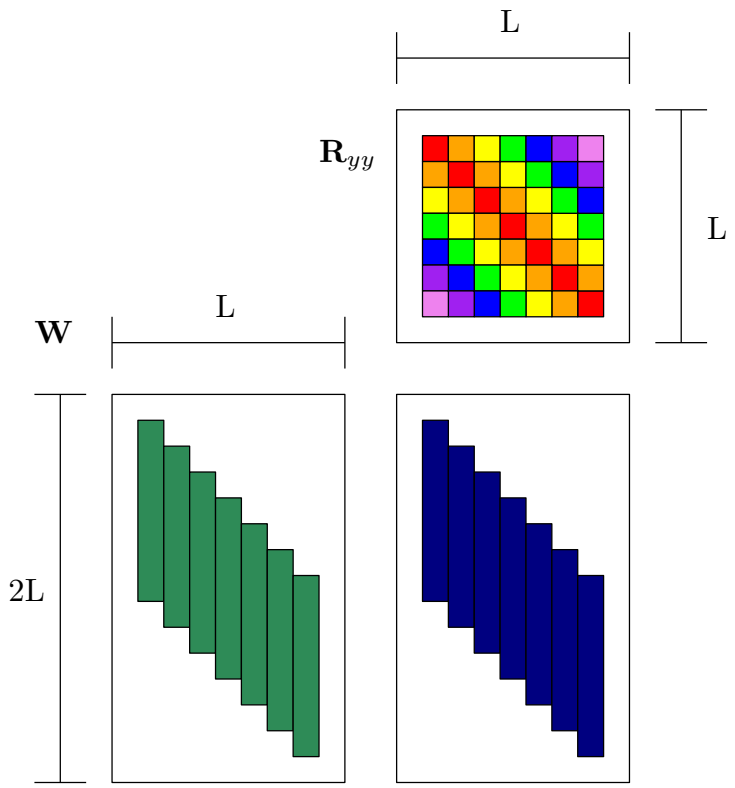

Figure 7.1: Illustration of the reduction of complexity through the Sylvester structure of the filter updates. Each column of the output is identical, meaning only the first column needs to be computed.

mathematical operations of the second-order statistics TRINICON algorithm presented in Section 7.2.2 are examples of parallelisable operations, with varying degrees of complexity. The elements in the correlation function vectors (Equations 7.6, 7.8 and 7.9) are independent from one another, i.e., frequency bin 1 of $\mathbf{r}_{xx_{11}}$ has no relation to frequency bin 2, allowing the element-wise

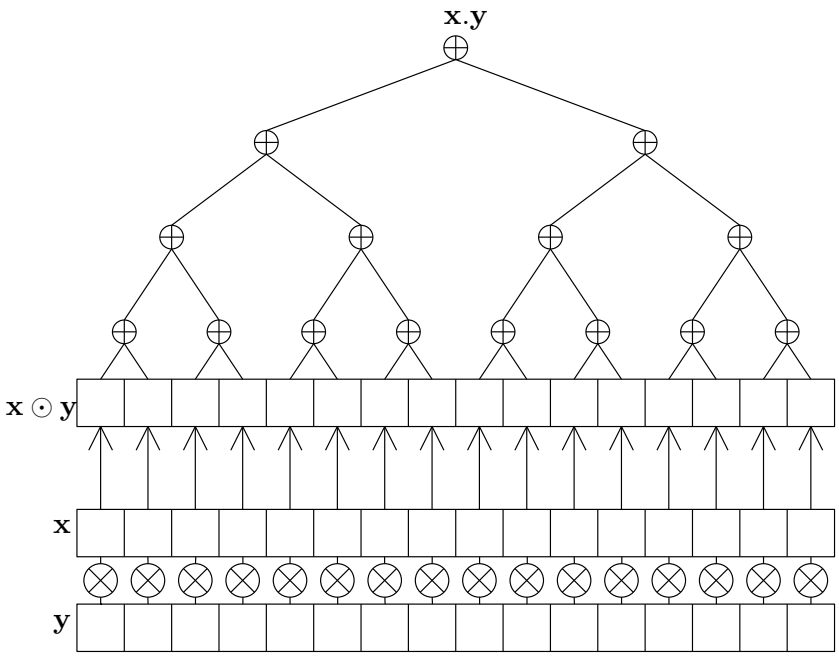

Figure 7.2: Structure of the parallel inner-product, an example of a divide-and-conquer algorithm

multiplication (and addition) to occur in separate operating threads without leading to synchronisation issues typical to multi-threaded programming. Similarly, the filter update operation (7.2) and the convolution operation to calculate the output channels (7.13) can be multi-threaded easily due to the same independence property. Computing $\mathbf{r}_{yy_{ii}}^{-1}$ (7.11) requires implementing the vector inner product which implies accumulating N intermediate element-wise products into a single output — leading to synchronisation and memory contention issues. The operation can be parallelised by reformulating the algorithm as a divide-and-conquer binary tree (Figure 7.2), which allows pairs of elements to be summed in parallel.

Current GPUs favour single-precision floating-point arithmetic over double-

precision, particularly on consumer-level GPUs, due to the costs associated with dedicated double-precision hardware. The single-precision performance, in terms of execution time, of GPUs can be up to 24 times greater than double-precision, which was the case with the GPU used in these investigations (NVIDIA GT650M). One of the objectives of this section was to investigate the viability of GPU processing for a large number of separation units using single-precision arithmetic for maximum efficiency. The resulting effects on separation accuracy were investigated to ensure that separation performance was not compromised significantly.

7.2.4 Simulation Setup

The implementation was evaluated by simulating two talkers in a highly reverberant three-dimensional room of dimensions 6.0m by 4.0m by 4.0m. The desired signal was located at (2.0m, 2.0m, 2.0m) and the interferer was located at (4.0m, 1.0m, 2.0m). The image source method [Allen and Berkley, 1979] was utilised to generate up to 4th order reflections from the 4 walls, with reflection coefficients set to 0.7. The two microphones were placed at (1.0m, 2.05m, 2.0m) and (1.0m, 1.95m, 2.0m), giving an array separation of 10cm. The block length (N) used in the TRINICON algorithm was set to 3072 samples and the separation filters (L) were set to 1024 taps. As input data, ten seconds of speech sampled at 16kHz were used. The input and filter vectors were zero padded to 4096 (the next nearest power of two) samples for efficient processing on the GPU. The regularisation parameter and weighting factors δ and ρ in (7.11) were set to 10^{-10} and 0.5 respectively. The online filter update forgetting factor in (7.12) was set to 0.25.

The computational performance was evaluated using a notebook computer

with an Intel i7-3635QM 4 core/8 thread CPU with a single threaded clock rate of up to 3.2GHz and an NVIDIA GT650M GPU with 384 threads running at a clock rate of 900MHz. The CPU reference implementation was single-threaded and implemented in both MATLAB using 64-bit floating-point arithmetic, as the accuracy reference, and native C code using 32-bit floating-point arithmetic using the fftw [FFTW, 2013] library to compute the FFT, as the computational performance reference.

7.2.5 Results

An unexpected result of the implementation was the increase in separation performance compared with the CPU-MATLAB reference code (Figure 7.3). As noted in Section 7.2.2, the implementation was coded using single-precision arithmetic, which was expected to adversely affect the separation performance by introducing greater rounding errors into the filter updates. The approximation used to compute $\mathbf{R}_{\mathrm{yy}_{\mathrm{ii}}}^{-1}$ in the filter update equation was suspected to be the source of the mismatch, involving the inverse of an inner product. After investigating the implementation carefully, the parallel inner-product was found to be the source of the discrepancy between the CPU-MATLAB and GPU separation performance. The serial implementation of the inner product can lead to significant rounding errors accumulating if there is a large dynamic range of values in the vectors. The parallel code by contrast, operates by summing pairs of values, which at each stage of computation, are independent from each other. As the pairs operate independently, the rounding errors are also independent, preventing a large rounding error in one pair swamping the other pairs. The result of this is that the overall rounding error in the sum is lower, leading to more accurate filters despite the lower

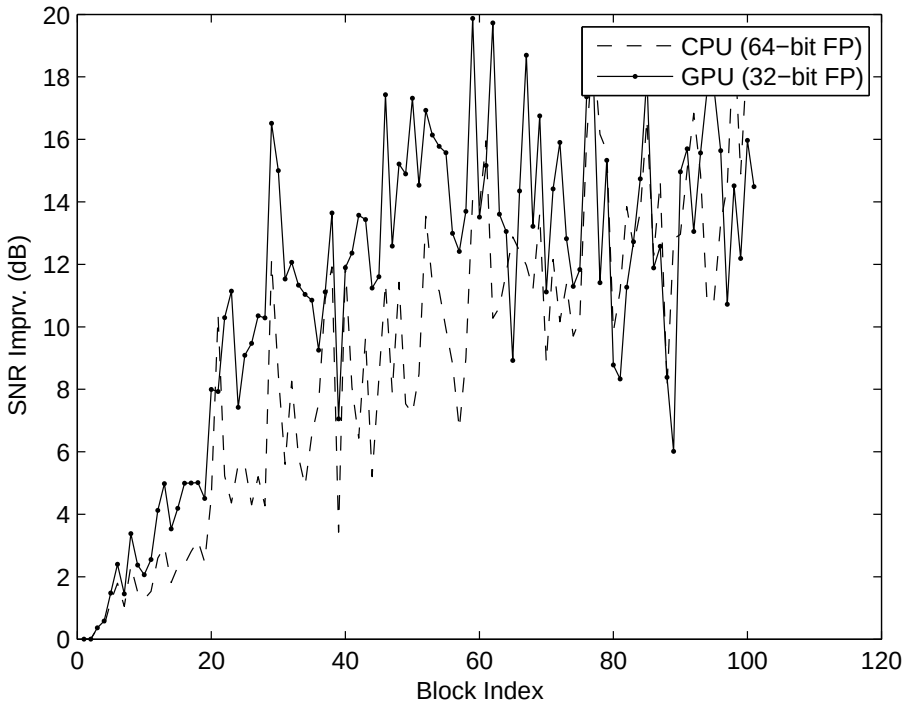

Figure 7.3: Signal to interferer ratio improvement

precision arithmetic. In [Yablonski, 2011] the author describes the effects of parallel algorithms and CPU/GPU instructions on accuracy of mathematical operations and demonstrates that the accuracy can be improved by exploiting parallel algorithms and/or special CPU/GPU instructions.

A modification to the CPU-MATLAB reference code verified the effect of the parallel inner product structure as the cause of the discrepancy (Figure 7.4). With the modification to the CPU code, the single-precision GPU implementation is comparable to the modified double-precision CPU-MATLAB code, differing only slightly in terms of both signal to noise power ratio improvement and signal distortion, which can be attributed to using lower

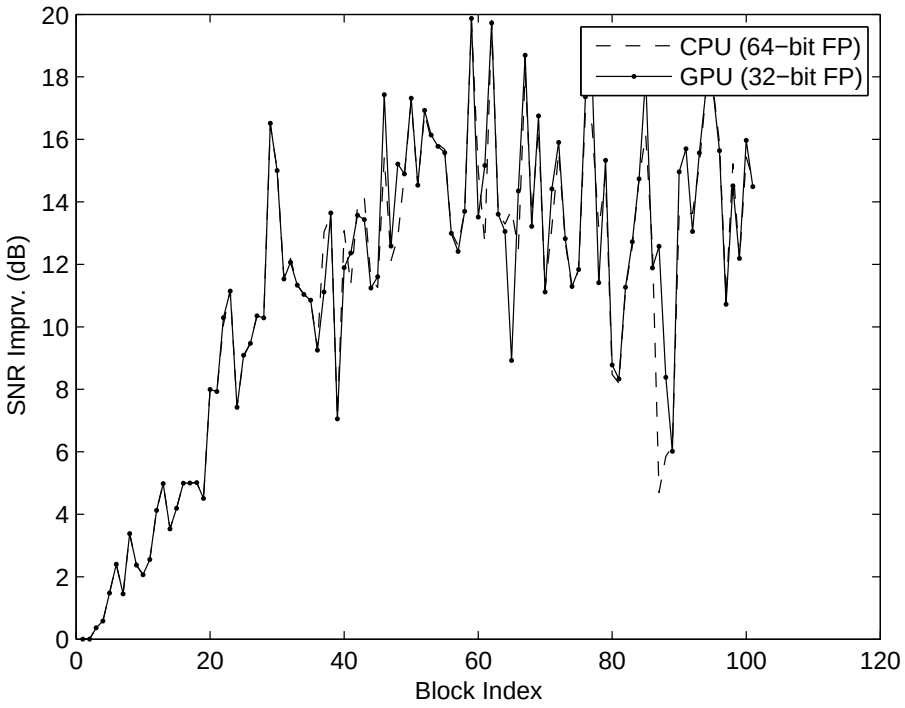

Figure 7.4: Signal to interferer ratio improvement with the CPU code modified to use a similar parallel structure for the dot-product

precision arithmetic. These modifications were tested using multiple data sets.

The signal distortion, the difference between the original undistorted speech and the desired speech after processing, closely matches the CPU-MATLAB implementation once the difference in inner product implementations are accounted for, as demonstrated in Figure 7.5. The computational performance of the GPU implementation is shown in Table 7.1. The simple two channel case, representative of a single-node in a distributed system, shows roughly an order of magnitude increase in performance over a C im-

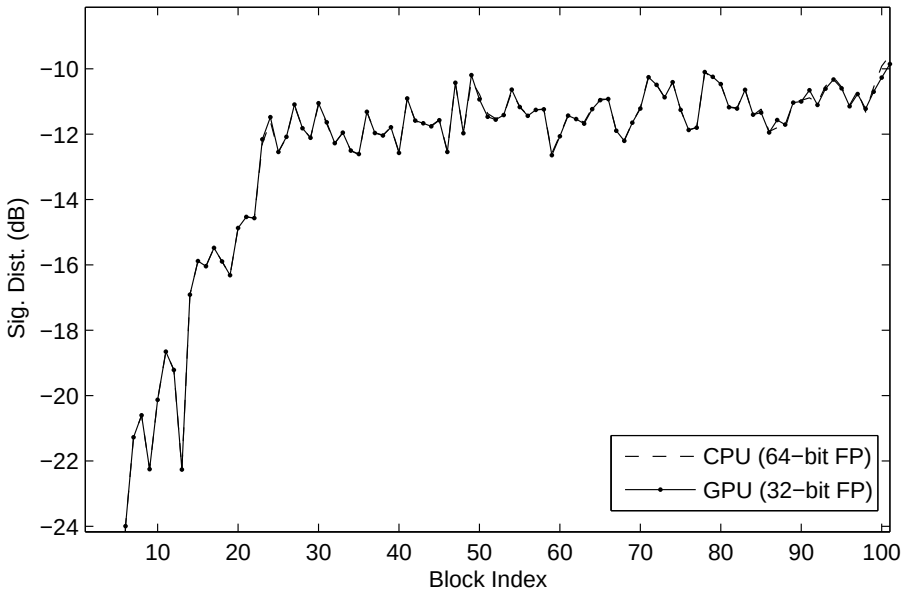

Figure 7.5: Signal distortion measure comparison between the CPU and GPU implementations

plementation running on a CPU. The single-node case, however, does not show the full potential of the GPU in accelerating the algorithm due to parallelisation limits in the FFT and inner product algorithms. A further simulation was conducted to evaluate the performance of the implementation operating on a large number of audio pairs to verify its potential in processing audio signals from a distributed set of microphones. A set of tests operating on 8, 16, 32 and 64 pairs of audio mixtures similar to the single demixer test were conducted.

The results presented in Table 7.2 show that additional gains in performance can be achieved by increasing the workload assigned to the GPU. Limitations of parallelism of the inner product and fast Fourier transform

	Time (ms)	Realtime (x)	Improv. (x)
CPU-MATLAB	2580	3.88	1
CPU-C	935	10.69	2.76
GPU	136	73.53	18.97

Table 7.1: Performance results for a single separation unit comparing single precision CPU-MATLAB, CPU-C and GPU implementations when separating a 10 second mixture.

Units	Time (ms)	Real-time (x)	Real-time/unit (x)
8	1262	23.78	190.24
16	2140	14.02	224.32
32	3898	7.70	246.40
64	7544	3.98	254.72

Table 7.2: Performance results for multiple simultaneous separation units

present themselves for small numbers of separation units. Both operations rely on a binary-tree decomposition as exhibited in Figure 7.2, which introduces two bottlenecks to the algorithm. The FFT and inner-product operations effectively reduce the number of operating threads to one in some portions of the algorithm. This limitation can be overcome by increasing the number of separation units to process, as each unit can be processed independently from one another. As seen in Table 7.2, the processing capability of the card appears to increase in terms of pairs per unit time, which demonstrates the advantages of processing multiple pairs of audio using the same device. The overall result shows that the card used in these simulations is theoretically

capable of processing more than 250 pairs in real-time, a substantial increase over the 10 pairs which can be processed using the single-threaded CPU implementation.

7.2.6 Discussion

The performance results demonstrated in Section 7.2.5 show the effectiveness of GPU acceleration for the TRINICON-BSS algorithm. Substantial increases in performance relative to the CPU-C implementation are attainable by offloading highly parallel portions of the algorithm to the GPU with counter-intuitively positive effects on the quality of the output compared with the double-precision CPU-MATLAB implementation. In addition, the GPU implementation is capable of processing a very large number of audio pairs in real-time on modest low-power hardware, demonstrating that it is potentially well suited as part of a quasi-distributive beamforming and blind source separation system. This GPU-accelerated algorithm is expected to be applicable to BSS-based signal extraction algorithms, such as the method described in [Reindl et al., 2013].

7.3 Real-time Robust Beamforming and BSS

7.3.1 Introduction

One of the primary motivations for this thesis was the issue of computational performance, particularly focussing on algorithms which could feasibly be implemented in real-time on a low power device. The beamforming methods covered in this thesis have been designed primarily for fixed-beamformer design, where the beamforming filters are designed offline and implemented as simple FIR filters with fixed coefficients. An adaptive system is constructed by using dual complementary fixed beamformers fed into an adaptive blind source separation system based on the TRINICON framework, which provides some ability to compensate for direction/position mismatch between the expected desired source location and the actual location. This implementation is a complete implementation of the algorithms described in Chapter 6.

7.3.2 System Design

The implementation consists of two stages, the first of which is the beamforming stage where the microphone inputs are filtered using the beam/null-formers designed using any of the methods described in previous chapters. The outputs of the beamforming stage are fed into a 2x2 second-order statistics implementation of the TRINICON framework, which was developed in the previous section.

In the beamforming stage, the samples are collected and windowed before transforming into the frequency domain. Each channel vector is then element-wise multiplied with the corresponding frequency domain beamforming filter

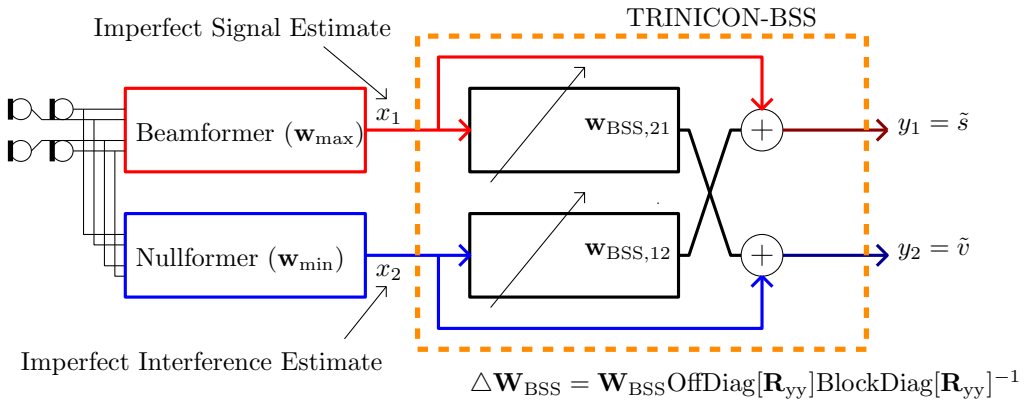

Figure 7.6: Overview of the implementation design.

to give the two intermediate inputs (the beam/null-formed data).

$$\mathbf{y}_1 = \sum_{n=1}^{N} \mathbf{w}_{\max,n} \circ \mathbf{x}_n \tag{7.14}$$

$$\mathbf{y}_2 = \sum_{n=1}^{N} \mathbf{w}_{\min,n} \circ \mathbf{x}_n \tag{7.15}$$

These intermediate inputs are then processed by the TRINICON stage to remove any mutual information between them, i.e., speech leaking into the nullformed channel, and/or interference leaking into the beamformed channel. The BSS filter updates are described in Chapter 6 (equations 7.6 through 7.12). Once the BSS filters have been updated for the current block, the signal and interference estimates are computed as

$$\tilde{s} = \mathbf{y}_1 + \mathbf{w}_{BSS,12} \circ \mathbf{y}_2 \tag{7.16}$$

$$\tilde{v} = \mathbf{y}_2 + \mathbf{w}_{BSS,21} \circ \mathbf{y}_1 \tag{7.17}$$

The signal and interference estimates are then transformed back into the time-domain and output through headphones/speakers or to disk.

The implementation was coded in C using compiler intrinsics to exploit the CPU single instruction multiple data (SIMD) vector instructions (AVX on Intel [Intel, 2015], and NEON on the ARM platforms [ARM, 2015b]) to accelerate vector addition, subtraction, multiplication, division, and summation operations. The AVX instructions allow the computation of 8 single precision floating-point operations per clock cycle, and the NEON instructions allow either 2 or 4 single precision floating-point operations per cycle (depending on the version of ARM processor).

7.3.3 Computational Performance

Computational performance was evaluated by processing pre-recorded 4-channel array signals through the program writing to disk rather than playing in real-time. The performance was evaluated by processing a pre-recorded demo multiple times to obtain a mean runtime for each processor tested. The objective was to find out whether the algorithm was feasible on low-power ARM devices with similar power to smartphones manufactured between 2012-2015. The three processors tested were a laptop processor: (Intel i7-3635QM, 3.2GHz peak clock rate); 2015 Raspberry Pi 2 [Raspberry Pi Foundation, 2015] (ARM Cortex-A7, 900MHz clock rate); and a 2015 ODROID-XU4 [Hard Kernel, 2015] (ARM Cortex-A15, 1.7GHz clock rate).

In Table 7.3 the mean performance times for the algorithm were compared for the various platforms and levels of processing. The objective was to achieve a real-time value of greater than 1 for the combined beamforming and BSS method, preferably much more to account for processor overhead associated with the operating system audio stack. The laptop could easily process the dual 4-channel beamformers and BSS system and provide glitch free audio

Table 7.3: Algorithm performance for the beamforming plus BSS system (and only beamforming or BSS) on various platforms using a 11.7 s length demonstration sampled at 44.1 kHz.

Processor	Time (s)	Realtime (x)
Intel i7-3635QM (3.2GHz)	0.43	27.21
2× 4-ch Beamforming	0.19	61.58
2-ch BSS	0.24	48.75
ARM Cortex A15 (1.7GHz)	2.38	4.91
2× 4-ch Beamforming	1.00	11.7
2-ch BSS	1.38	8.48
ARM Cortex A7 (900MHz)	10.01	1.17
2× 4-ch Beamforming	5.30	2.21
2-ch BSS	4.71	2.48

when played in real-time. Of the ARM processors, only the ODROID could manage real-time glitch free audio; the Raspberry Pi could process the data in real-time but was unable to play the audio glitch free, however it should be possible to use the algorithm at a lower sample rate. Currently the software is not well optimised (other than simple SIMD vector operations), nor is it multi-threaded to exploit CPU-core parallelism, and as such there may be significant processing improvements within reach. An interesting observation was that the 2-channel BSS system requires a comparable amount of processing power as the dual 4-channel beamformers. If the dual beamformers were extended to a full GSC implementation, i.e., 4 simultaneous 4-channel beamformers, the

processing requirements would be similar to the proposed system, assuming the adaptive filter after the GSC blocking matrix had negligible performance impact. A GSC implementation would likely require some kind of additional repair mechanism to ensure desired signal leakage was minimal, which would require additional processing steps.

7.3.4 Interference/Noise Reduction Performance

A simple demonstration of the implementation was conducted using a 4-element rectangular electret-microphone array, as laid out in Figure 7.7, fed into an XMOS USB Audio2 Interface multichannel sound card attached to a notebook computer. The speech source used was a 60-second example from the TIMIT [Garofolo et al., 1993] database played through a KEMAR head and torso simulator. The demonstration took place in a highly reverberant office/laboratory with significant levels of background interference — diffuse air conditioning, and with people talking and walking around. In addition, there was significant electrical noise present from the pre-amplifier sitting between the microphones and sound card. The microphones were not calibrated, and the flexible wire mounting resulted in errors in the measured positions of the microphones used to design the beamformer filters, i.e., moving the array tended to slightly shift the microphone positions relative to each other.

The three beamforming techniques developed in Chapters 4 and 5 were tested using the rectangular array. The 3D far-field beamformer was designed using (4.16) and setting the von Mises-Fisher coefficient $\kappa = 500$; the near-field beamformer was designed using the radial Gaussian method (4.82) with $\sigma = 0.04\,\mathrm{m}$; and the scatterer-based design used a head radius of $r_\mathrm{h} = 0.0875\,\mathrm{m}$, and a robust source design assuming a von Mises-Fisher distribution of sources

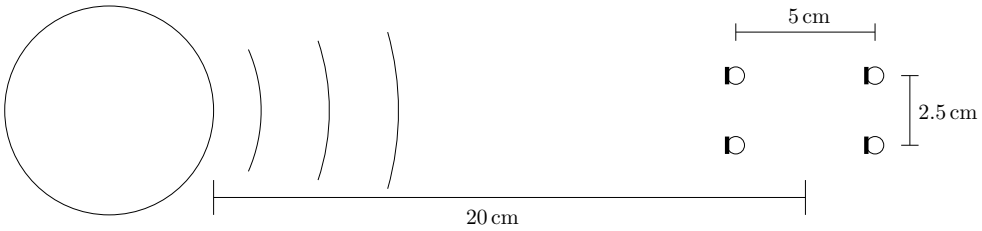

Figure 7.7: Layout of the rectangular array relative to the head and torso simulator.

on the head with a distribution parameter of $\kappa = 500$. The microphone array was placed 20 cm from the mouth of the head and torso simulator — although this is not far-field, the far-field beamformer design performed fairly similarly to the near-field/scatterer designs. The far-field designs are unable to exploit the wave amplitude differences between microphones in the array, but are able to use phase information to partially align signals. In practice, this meant that the beamforming channel performed similarly to the near-field/scatterer-based designs with only a small performance penalty in terms of SINR before post-processing. The far-field nullformer channel however did behave significantly differently compared with the near-field/scatterer-based designs, and as a result did not optimally collect background interference. The beamformer filter length was set to 1024-taps with a low-pass cut-off set to approximately 6 kHz to prevent spatial aliasing issues from occurring — the interference/noise level above 6 kHz was too little for the low-pass filter in the beamformer to have a significant interference/noise reduction effect on its own. The BSS configuration used was as in Section 6.2.3. The recording sample rate was

44.1 kHz. σ_n^2 used to design the sensor noise spatial correlation matrix was set to 10^{-3} to simultaneously tolerate the array imperfections detailed above, and to improve the eigenvalue/vector solution (2.102) stability.

In Figure 7.8, the waveforms for the raw microphone data (the highest SINR channel is presented), far-field design output, near-field design output, and scatterer-design output are presented. All three beamforming methods resulted in significant improvement in SINR, which improved from the unprocessed value of between 10-15 dB during the speech utterances to between 25-30 dB after processing using beamforming and BSS (Table 7.4). In Figure 7.9 a small subset of data is shown. It can be seen that all three beamformer designs are extremely effective at filtering the low-mid frequency noise (below 1.5 kHz). The far-field design had a small issue with a band of noise present at about 4.2 kHz which is successfully filtered by the near-field and scatterer-based designs, this was due to the signals received in the far-field nullformer channel — which did not include the band of noise compared with the other designs and thus was not filtered out of the primary channel. This is due to suboptimal null design when applying the far-field design to a near-field problem. The near-field and scatterer-based designs perform nearly identically with no significant differences between the two techniques. As found in Chapter 5, this was not unexpected as the diffuse interference (air conditioner) was the dominant interference source, and as the simulations showed in Section 5.2.7, the performance of the scatterer-based design did not differ significantly from the free-field design under isotropic interference conditions.

The BSS post-processor performed as intended, removing the desired speech component from the nullformer channel, and removing both the tran-

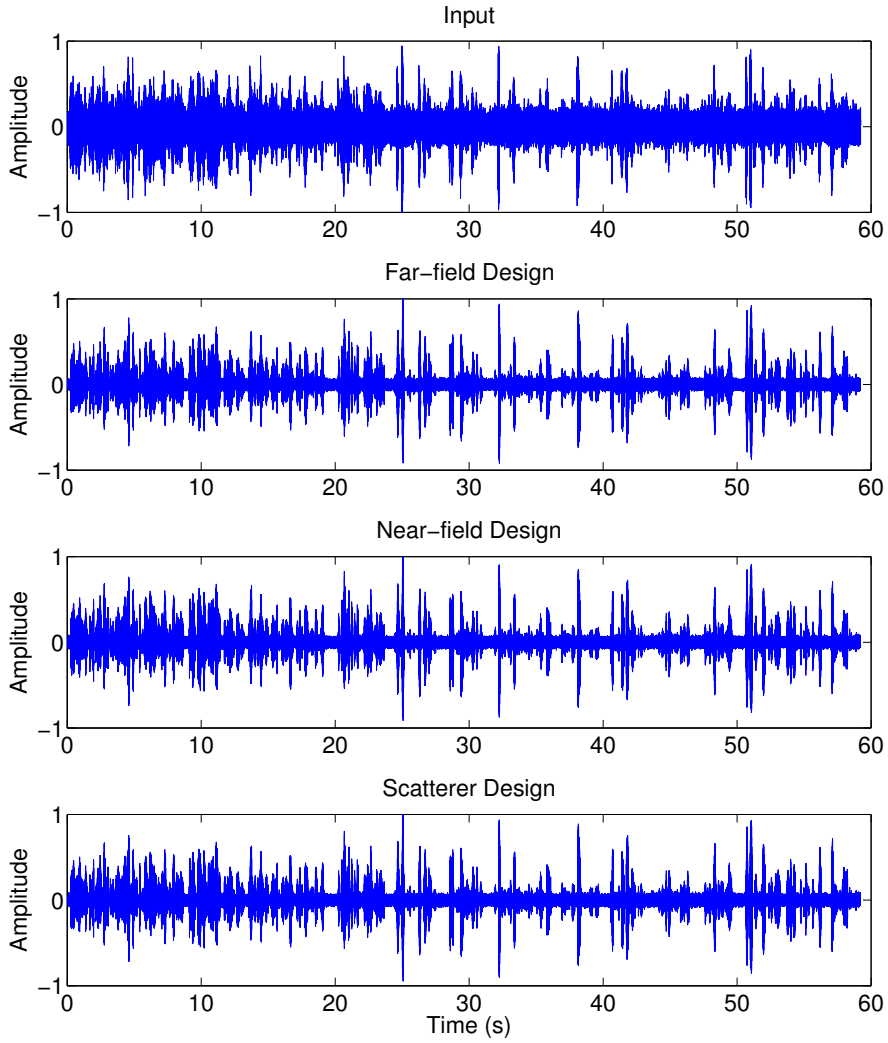

Figure 7.8: Comparison between the unprocessed microphone signal and the beamformer plus BSS processed output for a real array using the robust far-field, near-field, and scatterer beamformer designs.

Table 7.4: Peak SINR improvement for the speech in diffuse interference real-world example

Beamformer	Peak SINR Improvement (dB)
Far-field	10.25
Near-field	14.83
Scatterer	14.18

sient background footsteps/secondary talkers and some of the air conditioning noise from the primary beamformer channel. The post-processor provided a roughly 5 dB improvement in SINR over just using the beamformer.

7.3.5 Conclusions

In this section it was demonstrated that the algorithm developed in Chapter 6 can be feasibly implemented on a low-power device and run in real-time using a four-channel microphone array. The beamforming techniques developed in this thesis perform well in terms of interference/noise reduction in conjunction with the BSS-based post-processor when tested with an imperfect microphone array.

The simple proof-of-concept demonstration did not fully test the interference/noise reduction performance of the algorithm. Further testing with calibrated, properly positioned microphones (in a 3D configuration) in a controllable environment would be the obvious next step.

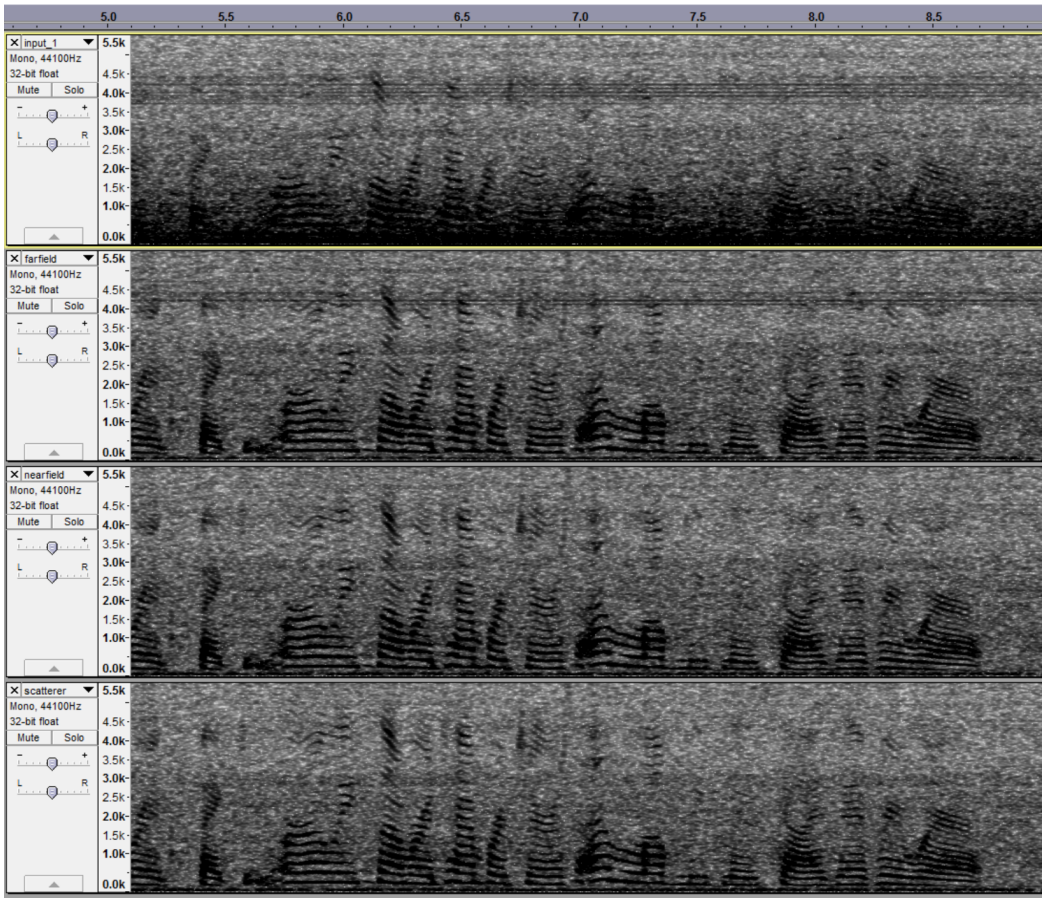

Figure 7.9: Short speech utterance from the demo data. From top to bottom: the raw input data (input channel with the best SINR); far-field beamformer; near-field beamformer; and scatterer-based beamformer. The spectrogram generated using Audacity [Audacity Team, 2015] used a Hamming window with a size of 2048 samples.

Chapter 8

Conclusions

8.1 Outline

This chapter highlights the conclusions and outlines a number of potential future research areas.

8.2 Conclusions

In Chapter 3, a robust beamformer formulation, using numerical integration to solve the correlation function, was used to derive a simple Wiener filtering technique. The robust method was quite effective for estimating approximate interference/noise statistics which led to an effective single channel Wiener filter solution. In Chapter 4, the robust beamforming technique was developed further by deriving a set of correlation functions for the specific applications of 2D/3D far-field, and 3D near-field beamforming. The far-field solutions derived reduced a computationally expensive integral solution to a pair of very simple analytic solutions for the correlation functions. The near-field solutions derived similarly reduced the complexity of the integral solution to a simpler summation solution. These correlation functions were found to be particularly useful in the design of robust nullformers used to estimate interference/noise signals.

In Chapter 5, a simple scattering model of sound propagation was considered in which the idealised point source was placed onto a solid sphere in order to approximate the effect a head has on wave propagation. It was found that this theoretical model delivers no significant improvements for highly reverberant scenarios, but may provide quite significant improvements for the application of directional interference blocking.

In Chapter 6, the dual beamformer and blind source separation algorithm was developed as a solution for interference/noise reduction in high noise environments. The method combined two fixed complementary beamformers designed to enhance and suppress a desired signal with an approximately known location relative to a multichannel microphone array. The blind

source separation system acted as a post-processor to minimise the mutual information between the beamformer outputs which provided some correction for errors in the assumptions used to design the front-end beamformers. The algorithm performs comparatively well against perfect information MVDR beamforming in diffuse noise, and can feasibly be implemented on a low-cost, low-power ARM platform, making it well suited for a number of mobile applications. The robust beamforming methods developed in Chapters 4 and 5 allow for a range of possible scenarios for far and near-field applications, including more realistic applications where scattering and diffraction are significant issues, which has been a neglected issue in the literature.

Including the BSS post-processor was intended to emulate the effect of a generalised sidelobe canceller design without signal leakage issues resulting from imperfect beamformer information (imperfect knowledge of the desired source location and/or interference/noise information). In Chapter 7, Section 2, it was observed that for the simple demonstration of the algorithm, all three beamformer designs (far-field, near-field, and scatterer) performed reasonably similarly when using the post-processor. This suggests that the BSS system is capable of refining the beamformers in order to improve SINR in the primary output channel, and improve the interference/noise reference in the secondary channel. Only limited testing occurred however, and there is scope to extend this work by testing the algorithm in a number of different interference/noise scenarios, different microphone array configurations and beamformer design parameters, and exploring BSS algorithms further.

8.2.1 Discussion

The correlation functions developed in Chapters 4 and 5 were found to be particularly useful for designing robust target suppressing nullformers. Existing techniques commonly used in the literature are extremely sensitive to even minor errors in the assumed source position knowledge and/or wave propagation model. The method presented in those chapters, while being technically sub-optimal, in practice may perform better for scenarios where target suppression is the objective.

Limited testing of the model developed in Chapter 5 showed that including scattering information produced no real benefit (nor downside) in terms of noise reduction compared with a near,free-field design. Designing beamformers using this method is computationally more expensive than the free-field designs, and combined with the minimal SINR performance advantage, suggests that this may not be a particularly useful design for many applications.

The proposed solution in Chapter 6 still requires an initial estimate of desired source position for the beamforming stage, and as such may not be suitable for applications where this is unavailable. The intended application for this solution is hand held devices in which the desired source position is expected to be approximately known. For other applications such as teleconferencing or hearing-aids, alternative strategies using direction of arrival estimation may be required in order for the algorithm to be useful.

8.2.2 Future Work

A number of avenues of future work have been identified which may improve on the performance of the algorithm developed during this thesis.

- A first step would be to further develop near-field prolate spheroid propagation models (such as those in [Barton et al., 2003]), which have also been used in the closely linked head-related transfer functions field [Jo et al., 2008] to model wave propagation around heads.

- This thesis has only focussed on idealised microphones (floating in free-space with no physical dimensions) and while this approximation worked surprisingly well during limited real-world testing, it obviously is not optimal. Modelling an enclosure (a simple rectangular box for example) using finite element methods [Reddy, 2006] could be a option to explore.

- Although voice activity detection/signal detection was identified as an issue, it would be expected that in the future, improved detection/recognition techniques (using video information for example [Joosten et al., 2015]) would decrease the probability of false positives/negatives in high noise environments. This would enable adaptive techniques such as those typically used to implement the MVDR/LCMV and GSC beamformers to identify desired source transfer function vectors and/or signal statistics more accurately and improve noise reduction as a result. In this case, the existing MVDR/LCMV and GSC techniques may become better solutions than the method proposed in this thesis, presuming the sensor fusion method was computationally efficient.

- The blind source separation component used a fixed value for the gradient ascent/descent filter adaptation (μ in (6.10)) which was selected as a conservative guess which happened to work for all of the scenarios tested. Ideally, this convergence parameter should be adaptive to allow rapid adaptation of filters for low SINR conditions, and as SINR conditions improve, decrease filter adaptation to ensure stability. In practice, it was difficult to design a reliable adaptive step-size mechanism using suggested methods from the literature (such as using covariance matrix eigenvalue information [Widrow and Stearns, 1985], or cost function minimisation measures [Aichner et al., 2005, (27)] for example) which produced stable filters in a variety of scenarios. Further investigations on optimal adaptive filter convergence could lead to improvements in BSS filter performance, as well as other adaptive filtering algorithms.

- The low-power implementation did not fully exploit the hardware available, and there is scope for simple algorithm optimisations which would improve performance significantly. For example, the FFT library used in the implementation (fftw) is not the fastest algorithm currently available on the ARM platform [Blake, 2012; ARM, 2015a]. Additionally, as part of this thesis, a GPU accelerated implementation of the TRINICON-BSS system was developed which should be feasible to implement on 2015-era smartphones (or similar devices), which include increasingly powerful graphics hardware, programmable through OpenCL [Khronos, 2015], which are well suited for general signal processing algorithms. Shifting parts of the implementation onto the graphics core would very easily allow the algorithm to run on low-end hardware without a significant

impact on other processing tasks the device may be performing.

References

[Abramowitz and Stegun, 1964] Abramowitz, M. and Stegun, I. A. (1964). *Handbook of Mathematical Functions: with Formulas, Graphs, and Mathematical Tables*, volume 55. Courier Dover Publications.

[Aichner et al., 2006] Aichner, R., Buchner, H., and Kellermann, W. (2006). A Novel Normalization and Regularization Scheme for Broadband Convolutive Blind Source Separation. In Rosca, J., Erdogmus, D., Príncipe, J. C., and Haykin, S., editors, *Independent Component Analysis and Blind Signal Separation*, volume 3889 of *Lecture Notes in Computer Science*, pages 527–535. Springer Berlin Heidelberg.

[Aichner et al., 2005] Aichner, R., Buchner, H., Yan, F., and Kellermann, W. (2005). A Real-Time Blind Source Separation Scheme and its Application to Reverberant and Noisy Acoustic Environments. *Signal Processing*, 86:1260–1277.

[Allen and Berkley, 1979] Allen, J. B. and Berkley, D. A. (1979). Image Method for Efficiently Simulating Small-room Acoustics. *The Journal of the Acoustical Society of America*, 65:943.

[Anderson, 2012] Anderson, C. A. (2012). Speech Enhancement using Multiple Transducers. Master's thesis, Victoria University of Wellington.

[Anderson et al., 2014a] Anderson, C. A., Meier, S., Kellermann, W., Teal, P. D., and Poletti, M. A. (2014a). A GPU-Accelerated Real-Time Implementation of TRINICON-BSS for Multiple Separation Units. In *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2014 4th Joint Workshop on*, pages 102–106.

[Anderson et al., 2015a] Anderson, C. A., Meier, S., Kellermann, W., Teal, P. D., and Poletti, M. A. (2015a). TRINICON-BSS System Incorporating Robust Dual Beamformers for Noise Reduction. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 529–533. IEEE.

[Anderson et al., 2014b] Anderson, C. A., Teal, P. D., and Poletti, M. A. (2014b). Multichannel Wiener Filter Estimation using Source Location Knowledge for Speech Enhancement. In *Statistical Signal Processing (SSP), 2014 IEEE Workshop on*, pages 57–60. IEEE.

[Anderson et al., 2015b] Anderson, C. A., Teal, P. D., and Poletti, M. A. (2015b). Spatially Robust Far-field Beamforming Using the von Mises(-Fisher) Distribution. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 23(12):2189–2197.

[ARM, 2015a] ARM (2015a). Ne10 FFT feature. http://community.arm.com/groups/android-community/blog/2013/12/18/projectne10-fft-is-updated.

[ARM, 2015b] ARM (2015b). NEON — ARM. http://www.arm.com/products/processors/ technologies/neon.php.

[Audacity Team, 2015] Audacity Team (2015). Audacity. http://audacityteam.org/.

[Ba et al., 2007] Ba, D. E., Florêncio, D., and Zhang, C. (2007). Enhanced MVDR Beamforming for Arrays of Directional Microphones. In *Multimedia and Expo, 2007 IEEE International Conference on*, pages 1307–1310. IEEE.

[Barton et al., 2003] Barton, J. P., Wolff, N. L., Zhang, H., and Tarawneh, C. (2003). Near-field Calculations for a Rigid Spheroid with an Arbitrary Incident Acoustic Field. *The Journal of the Acoustical Society of America*, 113(3).

[Benesty et al., 2008] Benesty, J., Chen, J., and Huang, Y. (2008). *Microphone Array Signal Processing*, volume 1. Springer Science & Business Media.

[Benesty et al., 2007] Benesty, J., Chen, J., Huang, Y. A., and Dmochowski, J. (2007). On Microphone-array Beamforming from a MIMO Acoustic Signal Processing Perspective. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(3):1053–1065.

[Blake, 2012] Blake, A. (2012). *Computing the Fast Fourier Transform on SIMD Microprocessors*. PhD thesis, University of Waikato.

[Boll, 1979] Boll, S. F. (1979). Suppression of Acoustic Noise in Speech using Spectral Subtraction. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 27(2):113–120.

[Buchner et al., 2004a] Buchner, H., Aichner, R., and Kellermann, W. (2004a). Blind Source Separation for Convolutive Mixtures: A Unified Treatment. In Benesty, J. and Huang, Y., editors, *Audio Signal Processing*, chapter 10, pages 255–293. Kluwer Academic Publishers.

[Buchner et al., 2004b] Buchner, H., Aichner, R., and Kellermann, W. (2004b). TRINICON: A Versatile Framework for Multichannel Blind Signal Processing. In *in Proc. IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, pages 889–892.

[Buckley and Griffiths, 1986] Buckley, K. M. and Griffiths, L. J. (1986). An Adaptive Generalized Sidelobe Canceller with Derivative Constraints. *Antennas and Propagation, IEEE Transactions on*, 34(3):311–319.

[Capon, 1969] Capon, J. (1969). High-Resolution Frequency-Wavenumber Spectrum Analysis. *Proceedings of the IEEE*, 57(8):1408–1418.

[Cardoso and Souloumiac, 1993] Cardoso, J.-F. and Souloumiac, A. (1993). Blind Beamforming for Non-Gaussian Signals. In *IEE Proceedings F (Radar and Signal Processing)*, volume 140, pages 362–370. IET.

[Catic et al., 2010] Catic, J., Dau, T., Buchholz, J. M., and Gran, F. (2010). The Effect of a Voice Activity Detector on the Speech Enhancement Performance of the Binaural Multichannel Wiener Filter. *EURASIP Journal on Audio, Speech, and Music Processing*, 2010(1):840294.

[Cauchi et al., 2014] Cauchi, B., Kodrasi, I., Rehr, R., Gerlach, S., Jukic, A., Gerkmann, T., Doclo, S., and Goetze, S. (2014). Joint Dereverberation and Noise Reduction using Beamforming and a Single-channel Speech

Enhancement Scheme. *Reverb Challenge. IEEE Audio, Acoust., Signal Process. TC*.

[Chen and Benesty, 2011] Chen, J. and Benesty, J. (2011). A Time-Domain Widely Linear MVDR Filter for Binaural Noise Reduction. In *Applications of Signal Processing to Audio and Acoustics (WASPAA), 2011 IEEE Workshop on*, pages 105–108. IEEE.

[Chen and Benesty, 2013] Chen, J. and Benesty, J. (2013). On the Time-Domain Widely Linear LCMV Filter for Noise Reduction with a Stereo System. *Audio, Speech, and Language Processing, IEEE Transactions on*, 21(7):1343–1354.

[Clapp, 1970] Clapp, R. E. (1970). Generalized Addition Theorem for Spherical Harmonics. *Journal of Mathematical Physics*, 11(1):1–4.

[Cohen, 2004] Cohen, I. (2004). Relative Transfer Function Identification using Speech Signals. *Speech and Audio Processing, IEEE Transactions on*, 12(5):451–459.

[Colton and Kress, 1998] Colton, D. and Kress, R. (1998). *Inverse Acoustic and Electromagnetic Scattering Theory*, pages 27–30. Springer.

[Cox et al., 1986] Cox, H., Zeskind, R., and Kooij, T. (1986). Practical Supergain. *Acoustics, Speech and Signal Processing, IEEE Transactions on*, 34(3):393–398.

[Dam et al., 2004] Dam, H. Q., Low, S. Y., Dam, H. H., and Nordholm, S. (2004). Space Constrained Beamforming with Source PSD Updates. In

Acoustics, Speech, and Signal Processing, 2004. Proceedings.(ICASSP'04). IEEE International Conference on, volume 4, pages iv–93. IEEE.

[Davis et al., 2005] Davis, A., Low, S. Y., Nordholm, S., and Grbic, N. (2005). A Subband Space Constrained Beamformer Incorporating Voice Activity Detection. In *Acoustics, Speech, and Signal Processing, 2005. Proceedings.(ICASSP'05). IEEE International Conference on*, volume 3, pages iii–65. IEEE.

[Doclo and Moonen, 2002] Doclo, S. and Moonen, M. (2002). GSVD-based Optimal Filtering for Single and Multimicrophone Speech Enhancement. *Signal Processing, IEEE Transactions on*, 50(9):2230–2244.

[Duda et al., 1999] Duda, R. O., Avendano, C., and Algazi, V. R. (1999). An Adaptable Ellipsoidal Head Model for the Interaural Time Difference. In *Acoustics, Speech, and Signal Processing, 1999. Proceedings., 1999 IEEE International Conference on*, volume 2, pages 965–968. IEEE.

[Fancourt and Parra, 2001] Fancourt, C. L. and Parra, L. (2001). The Coherence Function in Blind Source Separation of Convolutive Mixtures of Non-stationary Signals. In *Neural Networks for Signal Processing XI, 2001. Proceedings of the 2001 IEEE Signal Processing Society Workshop*, pages 303–312. IEEE.

[FFTW, 2013] FFTW (2013). FFTW Home Page. http://www.fftw.org/.

[Fisher, 1953] Fisher, R. (1953). Dispersion on a Sphere. *Proceedings of the Royal Society of London. Series A. Mathematical and Physical Sciences*, 217(1130):295–305.

[Flanagan et al., 1991a] Flanagan, J., Berkley, D., Elko, G., West, J., and Sondhi, M. (1991a). Autodirective Microphone Systems. *Acta Acustica united with Acustica*, 73(2):58–71.

[Flanagan et al., 1985] Flanagan, J., Johnston, J., Zahn, R., and Elko, G. (1985). Computer-steered Microphone Arrays for Sound Transduction in Large Rooms. *The Journal of the Acoustical Society of America*, 78(5):1508–1518.

[Flanagan et al., 1991b] Flanagan, J. L., Mammone, R., and Elko, G. W. (1991b). Autodirective Microphone Systems for Natural Communication with Speech Recognizers. In *Proceedings of DARPA Speech and Natural Language Workshop*, pages 170–175.

[Foshati and Khunjush, 2013] Foshati, A. and Khunjush, F. (2013). A Novel Implementation of Double Precision and Real Valued ICA Algorithm for Bioinformatics Applications on GPUs. In *Euro-Par 2012: Parallel Processing Workshops*, volume 7640 of *Lecture Notes in Computer Science*, pages 285–294. Springer Berlin Heidelberg.

[Frost, 1972] Frost, O.L., I. (1972). An Algorithm for Linearly Constrained Adaptive Array Processing. *Proceedings of the IEEE*, 60(8):926–935.

[Gannot et al., 2001] Gannot, S., Burshtein, D., and Weinstein, E. (2001). Signal Enhancement Using Beamforming and Nonstationarity with Applications to Speech. *IEEE Trans. Signal Processing*, (49):1614–1626.

[Gannot and Cohen, 2002] Gannot, S. and Cohen, I. (2002). Speech Enhancement Based on the General Transfer Function GSC and Postfiltering. In *IEEE Trans. Speech and Audio Processing*.

[Garofolo et al., 1993] Garofolo, J. S., Lamel, L. F., Fisher, W. M., Fiscus, J. G., and Pallett, D. S. (1993). DARPA TIMIT acoustic-phonetic continous speech corpus CD-ROM. NIST speech disc 1-1.1. *NASA STI/Recon Technical Report N*, 93:27403.

[Gradshteyn and Ryzhik, 2007] Gradshteyn, I. S. and Ryzhik, I. M. (2007). *Table of Integrals, Series, and Products*. Table of Integrals, Series, and Products Series. Elsevier Science.

[Grbic et al., 2003] Grbic, N., Nordholm, S., and Cantoni, A. (2003). Optimal FIR Subband Beamforming for Speech Enhancement in Multipath Environments. *IEEE Signal Processing Letters*, 10(11):335–338.

[Griffiths and Jim, 1982] Griffiths, L. and Jim, C. (1982). An Alternative Approach to Linearly Constrained Adaptive Beamforming. *IEEE Transactions on Antennas and Propagation*, 30:27–34.

[Habets et al., 2009] Habets, E. A. P., Benesty, J., Gannot, S., Naylor, P. A., and Cohen, I. (2009). On the Application of the LCMV Beamformer to Speech Enhancement. In *Applications of Signal Processing to Audio and Acoustics, 2009. WASPAA'09. IEEE Workshop on*, pages 141–144. IEEE.

[Hager, 1989] Hager, W. W. (1989). Updating the Inverse of a Matrix. *SIAM Review*, 31(2):pp. 221–239.

[Hard Kernel, 2015] Hard Kernel (2015). ODROID-XU4. http://www.hardkernel.com/main/ main.php.

[Haykin, 1991] Haykin, S. (1991). *Adaptive Filter Theory*, pages 396–399. Prentice Hall.

[Haykin and Chen, 2005] Haykin, S. and Chen, Z. (2005). The Cocktail Party Problem. *Neural computation*, 17(9):1875–1902.

[Herbordt and Kellermann, 2001] Herbordt, W. and Kellermann, W. (2001). Computationally Efficient Frequency-Domain Robust Generalized Sidelobe Canceller. In *Proc. Int. Workshop on Acoustic Echo and Noise Control*, pages 51–55.

[Hoshuyama et al., 1999] Hoshuyama, O., Sugiyama, A., and Hirano, A. (1999). A Robust Adaptive Beamformer with a Blocking Matrix using Coefficient-Constrained Adaptive Filters. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, 82(4):640–647.

[Huopaniemi et al., 1999] Huopaniemi, J., Kettunen, K., and Rahkonen, J. (1999). Measurement and Modeling Techniques for Directional Sound Radiation from the Mouth. In *Applications of Signal Processing to Audio and Acoustics, 1999 IEEE Workshop on*, pages 183–186. IEEE.

[Hyvärinen, 2001] Hyvärinen, A. (2001). Fast ICA by a fixed-point algorithm that maximizes non-Gaussianity. *Independent Component Analysis: Principles and Practice*, page 71.

[Hyvärinen et al., 2004] Hyvärinen, A., Karhunen, J., and Oja, E. (2004). *Independent Component Analysis*, volume 46. John Wiley & Sons.

[Hyvärinen and Oja, 1997] Hyvärinen, A. and Oja, E. (1997). A fast fixed-point algorithm for independent component analysis. *Neural computation*, 9(7):1483–1492.

[Intel, 2015] Intel (2015). ISA Extensions Intel AVX. https://software.intel.com/en-us/isa-extensions/intel-avx.

[ITU-T, 2002] ITU-T (2002). Perceptual evaluation of speech quality (PESQ): An objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs. Technical report, ITU.

[Jeub and Vary, 2010] Jeub, M. and Vary, P. (2010). Binaural Dereverberation based on a Dual-Channel Wiener Filter with Optimized Noise Field Coherence. In *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on*, pages 4710–4713. IEEE.

[Jo et al., 2008] Jo, H., Park, Y., and Park, Y.-s. (2008). Approximation of Head Related Transfer Function using Prolate Spheroidal Head Model. *International Congress on Sound and Vibration, Proceedings of the*, pages 2963–2970.

[Jolliffe, 2002] Jolliffe, I. (2002). *Principal Component Analysis*. Wiley Online Library.

[Joosten et al., 2015] Joosten, B., Postma, E., and Krahmer, E. (2015). Voice Activity Detection Based on Facial Movement. *Journal on Multimodal User Interfaces*, pages 1–11.

[Kellermann et al., 2006] Kellermann, W., Buchner, H., and Aichner, R. (2006). Separating Convolutive Mixtures with TRINICON. In *Acoustics, Speech and Signal Processing, 2006. ICASSP 2006 Proceedings. 2006 IEEE International Conference on*, volume 5, pages V–V. IEEE.

[Khronos, 2015] Khronos (2015). The open standard for parallel programming of heterogeneous systems. https://www.khronos.org/opencl/.

[Kumatani et al., 2007] Kumatani, K., Gehrig, T., Mayer, U., Stoimenov, E., McDonough, J., and Wolfel, M. (2007). Adaptive Beamforming With a Minimum Mutual Information Criterion. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(8):2527–2541.

[Laugesen et al., 2003] Laugesen, S., Rasmussen, K. B., and Christiansen, T. (2003). Design of a Microphone Array for Headsets. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 37–40. IEEE.

[Li and Duraiswami, 2007] Li, Z. and Duraiswami, R. (2007). Flexible and Optimal Design of Spherical Microphone Arrays for Beamforming. *Audio, Speech, and Language Processing, IEEE Transactions on*, 15(2):702–714.

[Lorenz and Boyd, 2005] Lorenz, R. G. and Boyd, S. P. (2005). Robust Minimum Variance Beamforming. *Signal Processing, IEEE Transactions on*, 53(5):1684–1696.

[Mailloux, 2005] Mailloux, R. J. (2005). Phased Array Antenna Handbook. *Boston, MA: Artech House*.

[Mammasis and Stewart, 2010] Mammasis, K. and Stewart, R. W. (2010). Spherical Statistics and Spatial Correlation for Multielement Antenna Systems. *EURASIP Journal on Wireless Communications and Networking*, 2010(1):307265.

[Martinez et al., 2015] Martinez, J., Gaubitch, N., and Kleijn, W. B. (2015). A Robust Region-Based Near-field Beamformer. In *Acoustics, Speech and Signal Processing (ICASSP), 2015 IEEE International Conference on*, pages 2494–2498. IEEE.

[Mazur and Mertins, 2011] Mazur, R. and Mertins, A. (2011). A CUDA Implementation of Independent Component Analysis in the Time-Frequency Domain. In *Proc. European Signal Processing Conference*, Barcelona, Spain.

[McCowan and Bourlard, 2003] McCowan, I. A. and Bourlard, H. (2003). Microphone Array Post-Filter Based on Noise Field Coherence. *IEEE Transactions on Speech and Audio Processing*, 11(6):709–716.

[Merks et al., 2014] Merks, I., Xu, B., and Zhang, T. (2014). Design of a High Order Binaural Microphone Array for Hearing Aids using a Rigid Spherical Model. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 3650–3654. IEEE.

[Moonen, 1993] Moonen, M. (1993). Systolic MVDR beamforming with Inverse Updating. In *Radar and Signal Processing, IEEE Proceedings for*, volume 140, pages 175–178. IET.

[Morse and Ingard, 1968] Morse, P. M. and Ingard, K. U. (1968). *Theoretical Acoustics*. Princeton University Press.

[Morse et al., 1948] Morse, P. M., Morse, P. M., and Morse, P. M. (1948). *Vibration and sound*, volume 2. McGraw-Hill New York.

[Nemes, 2010] Nemes, G. (2010). New Asymptotic Expansion for the Gamma Function. *Archiv der Mathematik*, 95(2):161–169.

[NVIDIA, 2013] NVIDIA (2013). CUDA Toolkit. https://developer.nvidia.com/cuda-toolkit.

[Parra and Alvino, 2002] Parra, L. and Alvino, C. (2002). Geometric Source Separation: Merging Convolutive Source Separation with Geometric Beamforming. *Speech and Audio Processing, IEEE Transactions on*, 10(6):352–362.

[Piersol, 1978] Piersol, A. (1978). Use of Coherence and Phase Data Between Two Receivers in Evaluation of Noise Environments. *Journal of Sound and Vibration*, 56(2):215–228.

[Rafaely, 2005] Rafaely, B. (2005). Analysis and Design of Spherical Microphone Arrays. *Speech and Audio Processing, IEEE Transactions on*, 13(1):135–143.

[Ramirez et al., 2004] Ramirez, J., Segura, J. C., Benitez, C., De La Torre, A., and Rubio, A. (2004). Efficient Voice Activity Detection Algorithms Using Long-term Speech Information. *Speech communication*, 42(3):271–287.

[Raspberry Pi Foundation, 2015] Raspberry Pi Foundation (2015). Raspberry Pi 2 Model B. https://www.raspberrypi.org/products/raspberry-pi-2-model-b/.

[Reddy, 2006] Reddy, J. (2006). *An Introduction to the Finite Element Method*. McGraw-Hill series in mechanical engineering. McGraw-Hill.

[Reindl et al., 2013] Reindl, K., Zheng, Y., Schwarz, A., Meier, S., Maas, R., Sehr, A., and Kellermann, W. (2013). A Stereophonic Acoustic Signal

Extraction Scheme for Noisy and Reverberant Environments. *Computer Speech and Language*, 27:726–745.

[Sawada et al., 2004] Sawada, H., Mukai, R., Araki, S., and Makino, S. (2004). A Robust and Precise Method for Solving the Permuation Problem of Frequency-Domain Blind Source Separation. *IEEE Transactions on Speech and Audio Processing*, 12(5):530–538.

[Schneider et al., 2012] Schneider, M., Schuh, F., and Kellermann, W. (2012). The Generalized Frequency-Domain Adaptive Filtering Algorithm Implemented on a GPU for Large-Scale Multichannel Acoustic Echo Cancellation. In *ITG Conference on Speech Communication*, pages 1–4.

[Schwarz and Kellermann, 2015] Schwarz, A. and Kellermann, W. (2015). Coherent-to-Diffuse Power Ratio Estimation for Dereverberation. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 23(6):1006–1018.

[Shabtai and Rafaely, 2014] Shabtai, N. R. and Rafaely, B. (2014). Generalized Spherical Array Beamforming for Binaural Speech Reproduction. *Audio, Speech, and Language Processing, IEEE/ACM Transactions on*, 22(1):238–247.

[Shahbazpanahi et al., 2003] Shahbazpanahi, S., Gershman, A. B., Luo, Z.-Q., and Wong, K. M. (2003). Robust Adaptive Beamforming for General-Rank Signal Models. *Signal Processing, IEEE Transactions on*, 51(9):2257–2269.

[Talmon et al., 2009] Talmon, R., Cohen, I., and Gannot, S. (2009). Relative Transfer Function Identification using Convolutive Transfer Function Ap-

proximation. *Audio, Speech, and Language Processing, IEEE Transactions on*, 17(4):546–555.

[Teal et al., 2002a] Teal, P. D., Abhayapala, T. D., and Kennedy, R. A. (2002a). Spatial Correlation for General Distributions of Scatterers. *Signal Processing Letters, IEEE*, 9(10):305–308.

[Teal et al., 2002b] Teal, P. D., Abhayapala, T. D., and Kennedy, R. A. (2002b). Spatial Correlation in Non-isotropic Scattering Scenarios. In *IEEE International Confererence on Acoustics Speech and Signal Processing*, volume 3, pages III–2833. IEEE.

[Van den Bogaert et al., 2009] Van den Bogaert, T., Doclo, S., Wouters, J., and Moonen, M. (2009). Speech Enhancement with Multichannel Wiener Filter Techniques in Multimicrophone Binaural Hearing Aids. *The Journal of the Acoustical Society of America*, 125(1):360–371.

[Van Trees, 2004] Van Trees, H. L. (2004). *Detection, Estimation, and Modulation Theory, Optimum Array Processing*. John Wiley & Sons.

[Ward and Elko, 1997] Ward, D. B. and Elko, G. W. (1997). Mixed Nearfield/Farfield Beamforming: a new Technique for Speech Acquisition in a Reverberant Environment. In *Applications of Signal Processing to Audio and Acoustics, 1997. 1997 IEEE ASSP Workshop on*, pages 4–pp. IEEE.

[Weninger et al., 2012] Weninger, F., Wöllmer, M., Geiger, J., Schuller, B., Gemmeke, J. F., Hurmalainen, A., Virtanen, T., and Rigoll, G. (2012). Non-negative Matrix Factorization for Highly Noise-robust ASR: to Enhance or

to Recognize? In *Acoustics, Speech and Signal Processing (ICASSP), 2012 IEEE International Conference on*, pages 4681–4684. IEEE.

[Whittaker and Watson, 1996] Whittaker, E. T. and Watson, G. N. (1996). *A Course of Modern Analysis*. Cambridge university press.

[Widrow and Stearns, 1985] Widrow, B. and Stearns, S. (1985). *Adaptive Signal Processing*. Prentice Hall.

[Williams, 1999] Williams, G. (1999). *Fourier Acoustics: Sound Radiation and Nearfield Acoustical Holography*. Academic Press.

[Wilson et al., 2008] Wilson, K. W., Raj, B., and Smaragdis, P. (2008). Regularized Non-negative Matrix Factorization with Temporal Dependencies for Speech Denoising. In *Interspeech*, pages 411–414.

[Yablonski, 2011] Yablonski, D. (2011). Numerical Accuracy Differences in CPU and GPGPU Codes. Master's thesis, Northeastern University.

[Yardibi et al., 2010] Yardibi, T., Bahr, C., Zawodny, N., Liu, F., III, L. C., and Li, J. (2010). Uncertainty Analysis of the Standard Delay-and-Sum Beamformer and Array Calibration. *Journal of Sound and Vibration*, 329(13):2654 – 2682.

[Zheng et al., 2009] Zheng, Y., Reindl, K., and Kellermann, W. (2009). BSS for Improved Interference Estimation for Blind Speech Signal Extraction with Two Microphones. *IEEE International Workshop on Computational Advances in Multi-Sensor Adaptive Processing*, pages 253–256.