



Circuits and Systems

Mekelweg 4,
2628 CD Delft
The Netherlands

<http://ens.ewi.tudelft.nl/>

CAS-2020-4815424

M.Sc. Thesis

Binaural beam-forming with dominant cue preservation for hearing aids

Vasudha Sathyapriyan

Abstract

For people with hearing impairment, it is important to have good speech intelligibility, while also being able to localise the sound sources. Many beam-forming algorithms for hearing aids have been proposed, that minimise the noise, in combination with spatial scene preservation of the target and the interferers. By constraining the spatial cues, the limited degrees of freedom (DoF) available for the design of the filter are expended, and this, to some extent degrades the noise reduction performance. Most of these methods try to preserve both the interaural time difference (ITD) and the interaural level difference (ILD) cues of the noise components over the entire frequency spectrum. However not all frequencies rely on both the ITDs and ILDs for the localisation of sound. More specifically, the ITDs are dominant cues in the frequencies $f < 1.5$ kHz and the ILDs are dominant cues in the frequencies $f \geq 1.5$ kHz. Based on these facts, in this thesis we try to preserve only the ILD cues of the noise components at frequencies above 1.5 kHz, while keeping the target signal undistorted. We investigate whether doing so saves the DoF that can be used to improve the noise reduction performance, in contrast to preserving both the cues. The thesis proposes two methods to preserve only the ILD cues of the interferers. The first method preserves the ILD cues perfectly, while the second method achieves a relaxed preservation of the ILD cues. Both methods show similar performance in anechoic and reverberant environments, and show that the noise reduction performance improves only mildly, when only the ILD cues of the noise components in the higher frequencies are preserved.



Binaural beam-forming with dominant cue preservation for hearing aids

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

ELECTRICAL ENGINEERING

by

Vasudha Sathyapriyan
born in Chennai, India

This work was performed in:

Circuits and Systems Group
Department of Microelectronics & Computer Engineering
Faculty of Electrical Engineering, Mathematics and Computer Science
Delft University of Technology



Delft University of Technology

Copyright © 2020 Circuits and Systems Group
All rights reserved.

DELFT UNIVERSITY OF TECHNOLOGY
DEPARTMENT OF
MICROELECTRONICS & COMPUTER ENGINEERING

The undersigned hereby certify that they have read and recommend to the Faculty of Electrical Engineering, Mathematics and Computer Science for acceptance a thesis entitled “**Binaural beam-forming with dominant cue preservation for hearing aids**” by **Vasudha Sathyapriyan** in partial fulfillment of the requirements for the degree of **Master of Science**.

Dated: 21-08-2020

Chairman:

dr.ir. R. C. Hendriks

Advisor:

ir. M. Çalış

Committee Members:

dr. J. C. Varon Perez

dr. S. Izadkhast

Abstract

For people with hearing impairment, it is important to have good speech intelligibility, while also being able to localise the sound sources. Many beam-forming algorithms for hearing aids have been proposed, that minimise the noise, in combination with spatial scene preservation of the target and the interferers. By constraining the spatial cues, the limited DoF available for the design of the filter are expended, and this, to some extent degrades the noise reduction performance. Most of these methods try to preserve both the ITD and the ILD cues of the noise components over the entire frequency spectrum. However not all frequencies rely on both the ITDs and ILDs for the localisation of sound. More specifically, the ITDs are dominant cues in the frequencies $f < 1.5$ kHz and the ILDs are dominant cues in the frequencies $f \geq 1.5$ kHz. Based on these facts, in this thesis we try to preserve only the ILD cues of the noise components at frequencies above 1.5 kHz, while keeping the target signal undistorted. We investigate whether doing so saves the DoF that can be used to improve the noise reduction performance, in contrast to preserving both the cues. The thesis proposes two methods to preserve only the ILD cues of the interferers. The first method preserves the ILD cues perfectly, while the second method achieves a relaxed preservation of the ILD cues. Both methods show similar performance in anechoic and reverberant environments, and show that the noise reduction performance improves only mildly, when only the ILD cues of the noise components in the higher frequencies are preserved.

Acknowledgments

This thesis would not have been possible without the help and support of many, who I would like to thank.

I would like to express my sincere gratitude to my supervisor dr.ir. R. C. Hendriks for his constant guidance and support. His tireless patience and valuable insights in answering my questions, and his confidence in me helped me progress in times of doubt.

Next, I would like to thank my daily advisor, ir. M. Çalış, for his help and support, especially during the last leg of my research. My special thanks go out to peers from the Msc. group, whose drive and enthusiasm towards their own projects motivated me to perform better.

Lastly, I would be remiss if I did not thank my family for their constant support and kind words, which kept me going in these uncertain times.

Vasudha Sathyapriyan
Delft, The Netherlands
21-08-2020

Contents

Abstract	v
Acknowledgments	vii
1 Introduction	1
1.1 Background	2
1.2 Research Question	4
1.3 Outline	4
2 Previous Work	5
2.1 Signal Model	5
2.2 Definitions	6
2.2.1 Cross power spectral density (CPSD)	6
2.2.2 Binaural Cues	7
2.2.3 Degrees of freedom (DoF)	8
2.3 Previous Work	8
2.3.1 Binaural minimum variance distortionless response (BMVDR)	8
2.3.2 Binaural linearly constrained minimum variance (BLCMV)	10
2.3.3 Joint binaural linearly constrained minimum variance (JBLCMV)	11
2.3.4 Relaxed binaural linearly constrained minimum variance (RBLCMV)	11
3 Methods Proposed	13
3.1 Interaural Level Difference Preservation Methods	13
3.1.1 Perfect interaural level difference cue preservation (P-ILD)	13
3.1.2 Relaxed interaural level difference cue preservation (R-ILD)	17
4 Results & Analysis	21
4.1 Experimental Setup	21
4.1.1 Acoustic Environment	21
4.1.2 Signal Synthesis	22
4.2 Performance Measures	23
4.2.1 Noise Reduction Measures	23
4.2.2 Intelligibility Measures	25
4.2.3 Localisation Measures	26
4.3 Simulation Results	26
4.3.1 Performance Analysis	27
4.4 Informal Listening Test	35
4.4.1 Test Setup	35
4.4.2 Results	36

5	Conclusion	41
5.1	Discussion	41
5.1.1	Method 1: P-ILD	41
5.1.2	Method 2: R-ILD	42
5.2	Recommendations	42
A	Mathematical Properties	45
A.1	Linear Algebra	45
A.2	Schur's Complement & Positive Definiteness	45
A.3	Lower bound of semi-definite relaxation (SDR)	45

List of Figures

4.1	Acoustic Environmental Setup.	22
4.2	Anechoic Environment: Comparing the target source localisation performance of the competing methods in terms of ILD, ITD and interaural transfer function (ITF).	27
4.3	Office Environment: Comparing the target source localisation performance of competing methods in terms of ILD, ITD and ITF.	27
4.4	Anechoic Environment: Comparing the localisation performance of the interfering sources amongst the competing methods in terms of ILD, ITD and ITF.	28
4.5	Office Environment: Comparing the localisation performance of the interfering sources amongst the competing methods in terms of ILD, ITD and ITF.	29
4.6	Comparing the noise reduction performance against the total ILD error over varying c (a) Anechoic, (b) Office.	31
4.7	Anechoic Environment: Comparing the noise reduction performance of the competing methods in terms of global segmental signal-to-noise ratio (gsSNR) and frequency weighted segmental signal-to-noise ratio (fwSegSNR).	32
4.8	Office Environment: Comparing the noise reduction performance of the competing methods in terms of gsSNR and fwSegSNR.	32
4.9	Anechoic Environment: Comparing the intelligibility performance of the competing methods in terms of short-term objective intelligibility (STOI) and speech intelligibility in bits (SIIB).	33
4.10	Office Environment: Comparing the intelligibility performance of the competing methods in terms of STOI and SIIB.	34
4.11	Comparing the noise reduction performance against the ILD localisation for the competing methods (a) Anechoic, (b) Office.	34
4.12	Anechoic Environment: Localisation error for each source.	38
4.13	Office Environment: Localisation error for each source.	38

List of Tables

4.1	Position of the Sources.	22
4.2	Anechoic Environment: Total Error in ILD over frequency bins.	30
4.3	Listening Test Scenario.	35
4.4	Two-way analysis of variance (ANOVA) test for Scene A.	36
4.5	Two-way ANOVA for Scene B.	37
4.6	T-test p-values for Scene A.	37
4.7	T-test p-values for Scene B.	37

Abbreviations

ANOVA Analysis Of Variance

ATF Acoustic Transfer Function

BLCMV Binaural Linearly Constrained Minimum Variance

BMVDR Binaural Minimum Variance Distortionless Response

BTE Behind-the-ear

CPSD Cross Power Spectral Density

DoF Degrees Of Freedom

FFT Fast Fourier Transform

fwSegSNR Frequency Weighted Segmental Signal-to-noise Ratio

gsSNR Global Segmental Signal-to-noise Ratio

HA Hearing Aid

HAD Hearing Assistive Device

HATS Head And Torso Simulator

HRIR Head Related Impulse Response

IC Interaural Coherence

ILD Interaural Level Difference

IPD Interaural Phase Difference

IR Impulse Response

ISTFT Inverse Short-time Fourier Transform

ITD Interaural Time Difference

ITF Interaural Transfer Function

JBLCMV Joint Binaural Linearly Constrained Minimum Variance

MMSE Minimum Mean Square Error

MVDR Minimum Variance Distortionless Response

MWF Multi-channel Wiener Filter

P-ILD Perfect Interaural Level Difference Cue Preservation

PSD Positive Semi-definite

QCQP Quadratically Constrained Quadratic Program
RBLCMV Relaxed Binaural Linearly Constrained Minimum Variance
R-ILD Relaxed Interaural Level Difference Cue Preservation
RLT Reformulation-linearisation Technique
SDP Semi-definite Program
SDR Semi-definite Relaxation
SIIB Speech Intelligibility In Bits
SNR Signal-to-noise Ratio
SRM Spatial Release From Masking
STFT Short-time Fourier Transform
STOI Short-term Objective Intelligibility
WGN White Gaussian Noise

The speech intelligibility of a listener gets degraded in the presence of background noise. This is true for those with healthy hearing as well as those suffering from hearing impairment. Since speech signals are redundant in nature, those with normal hearing may comprehend speech, despite the background noise. For the hearing impaired, however, it is sometimes already hard to understand the signal in a quiet scenario. This can be partly compensated by the use of hearing assistive devices (HADs) such as hearing aids (HAs), which use signal processing to reliably reduce noise and improve speech intelligibility [1].

Over the years, several single and multiple microphone strategies for noise reduction have been developed for HAs. These algorithms have grown to consider the attributes of the listener's auditory system, and the characteristics of the desired sound while enhancing the signals reaching the ears [2]. With the increased complexity of the auditory environment being processed, multi-microphone algorithms that exploit the spatial, in addition to the spectral and the temporal information, are preferred over the single channel strategies [3].

Commonly, modern HAs come in pairs, with multiple microphones in each HA unit. The signals from these microphones are combined, by properly delaying and summing the signals, to enhance the desired signal. This is commonly known as beam-forming. By applying complex weights to the signals in each microphone, and combining them, the signal from a desired direction can be enhanced [4].

Due to wireless communication, it has become possible to use the microphone signal measurements from both HA units. When each HA unit works independently to estimate the desired signal at each ear, the system is called bilateral HAs system. These HAs provide good noise reduction performance. They however generate monoaural outputs, that do not allow the user to localise sound, which would naturally occur, from the use of the signals at the two ears [5].

When the microphone signal measurements from both the HAs are used jointly, to estimate the filter weights that enhance the desired signal, it is called a binaural HAs system. With binaural hearing, comes the potential advantage of the improved ability to hear speech in noise, due to sound source localisation [6]. The importance of binaural hearing can be understood by the ability of a normal hearing person to localise and separate sound sources, which improves speech intelligibility due to spatial release from masking (SRM) [7]. SRM enables the listener to hear the desired sound better, as the noise and the target sources are spatially separated due to localisation. Using binaural beam-forming, thus helps to improve speech intelligibility by aiding the user

with sound localisation, in co-operation with noise reduction.

1.1 Background

There are several factors that affect how sound is perceived. The source location of a sound can be determined by the acoustic cues that are generated due to the difference in the path taken by the signals to reach the two ears, and from the interaction of sound with the head, torso and pinnae. These acoustic cues can be broadly categorised as monaural and binaural cues. The interaural time difference (ITD), the interaural level difference (ILD), and the interaural coherence (IC) cues, are determined by the difference in the signals reaching the left and the right ear, hence called binaural cues. While spectral cues, that are direction dependent and result in peaks and notches in the signal due to the pinnae effects at each ear, are called monaural cues [8]. The ITD spatial cues are rendered by the difference in the time taken for a signal to reach either ear from its source location. The ILD cues are due to the ‘head shadowing effect’, that is caused due to the reflections of the signal by the head, causing a drop in the energy of the signal reaching the ear away from the source [9]. Apart from these two cues that are based on the difference between signals, the IC cue compares the similarity between the signals that reach the two ears. Both the monaural and the binaural cues help with sound localisation. Psycho-physical experiments done in [10], however, suggest that the ILD and the ITD cues play a major role with sound localisation in the horizontal plane. Hence in this thesis, only the binaural cues, the ILD and the ITD, that help with localisation in the horizontal plane are considered.

According to the Duplex Theory of sound localisation, the ILD and the ITD cues operate in complementary frequency spectrum, for pure tone stimuli. The ITD cues are used in the lower frequencies below 1.5 kHz, while ILD cues are more pronounced in the higher frequencies for localisation [11]. This is because, in the higher frequencies, the wavelength of sound is shorter than the width of the head, causing the signal to reflect and lose energy as it reaches the ear away from the source. While in the lower frequencies, the wavelength of sound is comparable to the width of the head, making it bend around the head as it reaches the ear away from the source. On the contrary, the ITD cues are not reliable at higher frequencies, as the period of the sinusoids are shorter than the maximum interaural delay and hence, result in ambiguous phase leads and lags that make the ITD cues unreliable.

Furthermore, experiments in [10], [12], discuss about the weighting of these cues in the entire audible frequency spectrum. In [10], the dominance of the ITD cues over the ILD cues, in the lower frequencies less than 1.5 kHz, for sound localisation was experimentally proved. Psycho-physical experiments showed how listeners, when provided with contradicting cues of broadband stimuli, localised based on the ITD cues. Moreover in [12], experiments with low pass, high pass and wide band stimuli were conducted, to analyse the weight given to the binaural cues by the auditory brain for sound localisation. The results agreed with [10], while also establishing the dominance of the ILD cues over the contradicting ITD cues, at higher frequencies.

In their daily lives, it is important for the HA users to localise the sound reaching their ears, for example, in traffic. Most of the multi-channel noise reduction algorithms used in HAs, however, distort the spatial cues of sources. Currently there are two classifications of algorithms that preserve spatial cues in hearing aids [13]. The first category, consists of algorithms that apply an identical, real valued gain to the reference signals at both the ears, in order to maintain the interaural cues. These algorithms may not perform the best in terms of noise reduction, as they effectively work like single channel algorithms, that do not use the spatial characteristics of the environment. The second category of algorithms, follow the approach of optimising a cost function, that uses the spectro-temporal and the spatial characteristics. These algorithms can be further divided into spatio-temporal filtering algorithms and spatial filtering algorithms.

Spatio-temporal beam-forming, allows for distortions in the target signal after processing [14]. The binaural multi-channel Wiener filter (MWF) is a spatio-temporal beam-forming algorithm. Here, the minimum mean square error (MMSE) estimate of the reference microphone signals are generated at the output [15]. In [16], it has been proved that using binaural MWF, for a single speech source, the ILD and the ITD cues are preserved, while the spatial cues of the noise components are distorted. To allow the spatial cue preservation of the noise components, additional ITF or ITD or/and ILD terms are added to the cost function [17], [18], [19]. These algorithms, however, do not completely preserve the cues of the noise components, but offer a trade-off between noise reduction and spatial cue preservation.

With spatial beam-forming, the target signal is maintained reliably [20]. The binaural minimum variance distortionless response (BMVDR) beam-former is one such spatial filtering algorithm. It is a binaural extension of the minimum variance distortionless response (MVDR) beam-former, that achieves optimum noise reduction in the presence of background noise and interfering sources. Since the target is constrained to be undistorted, after processing, the spatial cues of the target signal are preserved. The interfering sources, however, after processing are co-located with the target, i.e., they take the spatial cues of the target [21]. By introducing additional constraints on the interferers, the binaural linearly constrained minimum variance (BLCMV) beam-former can preserve the spatial cues of a limited number of interferers [22].

The BLCMV formulation was further simplified in [23], which allowed for the spatial cue preservation of more interferers. Since the interferers may not be audible after beam-forming, in [24], the authors proposed a method where the cue preservation is applied only in those time-frequency tiles where the interfering sources are perceptually audible. Furthermore, in [25], a method to select the interfering sources, whose cues are to be preserved in algorithms that limit the number of interferers, was proposed. This is done by preserving the cues of those interferers that are perceptually audible at the output of the beam-former.

Nevertheless, the spatial cue preservation of the noise components were still heavily constrained by the equality constraints in the problem formulation of the joint binaural linearly constrained minimum variance (JBLCMV) beam-former [23]. In [26], a relaxation to the ITF cue preservation is introduced. As the inequality constraints replace the equality constraints of the JBLCMV formulation, more degrees of freedom (DoF) are available for the spatial cue preservation of more interferers. Moreover, for a fixed number of interferers, the noise reduction performance also improves, due to the increased DoF. By choosing a suitable allowable error in the ITF cues, a good trade-off between noise reduction and spatial cue preservation of the interferers was achieved.

In this thesis, we focus on spatial beam-forming, that preserves the target signal, in combination with spatial cue preservation of the interferers.

1.2 Research Question

The motivation of this thesis stems from two interesting observations, that

1. by relaxing the spatial cue preservation, for a given number of interferers, the noise reduction performance is better than JBLCMV [26].
2. binaural cues are frequency selective [10].

The methods discussed in section 1.1 preserve the spatial location of the noise components, by collectively preserving both, the ILD and the ITD cues. However, the ITD cues are dominant in the lower frequencies $f < 1.5$ kHz, and the ILD cues are dominant in the higher frequencies $f > 1.5$ kHz. This principle forms the foundation to this thesis, that intends to answer the following question.

- Will beam-forming with only the dominant binaural cue preservation of the noise components, help to improve the noise reduction performance, as opposed to the preservation of both the interaural time difference (ITD) and the interaural level difference (ILD) cues?

In this thesis, a first step is taken to answer the above question, by preserving only the ILD cues in the higher frequencies.

1.3 Outline

To address the above research question and to investigate a solution, the thesis is organised as follows. Chapter 2 introduces the signal model used, and follows up with a discussion on the binaural beam-forming algorithms available. Chapter 3 introduces the problem formulation, that is proposed to analyse the ILD cue preservation binaural beam-forming in the higher frequencies. Chapter 4, analyses the results from the simulations using the methods proposed, and compares the performance against the reference methods discussed in Chapter 2. Finally, Chapter 5 discusses a scope for further development of the work proposed, and concludes the work of this thesis.

This chapter describes the signal model used in the rest of the thesis. It discusses the methods that have been proposed previously to preserve the spatial cues in binaural beam-forming. The methods discussed here are later compared against the work proposed in this thesis.

2.1 Signal Model

Consider a binaural hearing aid configuration, having one HA on the left ear and one on the right ear, each with a microphone array containing $\frac{M}{2}$ microphones. The signals from the left HA are assumed to be transmitted wirelessly to the right HA, and vice versa, leading to a total of M signal measurements.

The signal measured in each microphone can be given as

$$y_m(t) = x_m(t) + n_m(t) + v_m(t) \quad m = 1, 2 \dots M, \quad (2.1)$$

with

y_m as the noisy signal received by the m^{th} microphone,
 x_m as the target source signal received by the m^{th} microphone,
 n_m as the interfering source signal received by the m^{th} microphone, and
 v_m as the background noise signal received by the m^{th} microphone.

It is assumed that there is one target signal with ‘ r ’ additive, mutually uncorrelated interfering signals, and uncorrelated noise.

For the ease of analysis, the signals are transformed to the frequency domain by the short-time Fourier transform (STFT).

$$y_m(l, k) = x_m(l, k) + n_m(l, k) + v_m(l, k), \quad m = 1, 2 \dots M, \quad (2.2)$$

where

l is the time frame index, and
 k is the frequency bin index.

As the processing is done independently per time frame, the time index ‘ l ’ is omitted for convenience. Altogether we have

$$\begin{aligned} y_m(k) &= x_m(k) + n_m(k) + v_m(k), \quad m = 1, 2 \dots M \\ &= a_m(k)s(k) + \sum_{i=1}^r b_{i,m}(k)u_i(k) + v_m(k), \end{aligned} \quad (2.3)$$

where

r is the number of interfering sources assumed,
 a_m is the acoustic transfer function (ATF) of the target signal to the m^{th} microphone,
 $b_{i,m}$ is ATF of the i^{th} interferer to the m^{th} microphone,
 s is the target signal at the target location, and
 u_i is the interference signal of the i^{th} interferer at its location.

Since the processing is done per frequency bin, the frequency bin index ‘ k ’ is omitted in the rest of the work for convenience. Taking the first $m = 1$ and the last $m = M$ microphone as the left and the right reference microphones respectively, all the measured signals can be combined into a single vector and can be represented as

$$\begin{aligned} \mathbf{y} &= [y_1 \ y_2 \ \cdots \ y_M]^T \in \mathbb{C}^{M \times 1} \\ &= \mathbf{a}s + \sum_{i=1}^r \mathbf{b}_i u_i + \mathbf{v} \\ &= \mathbf{x} + \mathbf{B}\mathbf{u} + \mathbf{v}, \end{aligned} \quad (2.4)$$

where

\mathbf{a} is the ATF vector of the target signal ($\mathbb{C}^{M \times 1}$),
 \mathbf{b}_i is the ATF vector of i^{th} interferer ($\mathbb{C}^{M \times 1}$),
 \mathbf{x} is the target signal vector ($\mathbb{C}^{M \times 1}$),
 \mathbf{u} is the interference signal vector ($\mathbb{C}^{r \times 1}$),
 \mathbf{v} is the noise signal vector ($\mathbb{C}^{M \times 1}$), and
 $\mathbf{B} = [\mathbf{b}_1 \ \mathbf{b}_2 \ \cdots \ \mathbf{b}_r] \in \mathbb{C}^{M \times r}$.

The spatial filtering algorithms that will be discussed, estimate the complex filtering coefficients of the binaural filter $\mathbf{w} = [\mathbf{w}_L^H \ \mathbf{w}_R^H]^H \in \mathbb{C}^{2M \times 1}$ such that, the output at the left and the right hearing aid can be given as

$$x_1 = \mathbf{w}_L^H \mathbf{y}, \quad x_M = \mathbf{w}_R^H \mathbf{y}. \quad (2.5)$$

2.2 Definitions

2.2.1 Cross power spectral density (CPSD)

As the target, interferers and the noise are mutually uncorrelated, the CPSD of the measured signal can be written as

$$\mathbf{P}_y = \mathbf{P}_x + \underbrace{\sum_{i=1}^r \mathbf{P}_{u_i}}_{\mathbf{P}} + \mathbf{P}_v, \quad (2.6)$$

where

$\mathbf{P}_y = \mathbf{E}\{\mathbf{y}\mathbf{y}^H\}$ is the CPSD matrix of the noisy signal ($\mathbb{C}^{M \times M}$),

$\mathbf{P}_x = \mathbf{E} \{ \mathbf{x}\mathbf{x}^H \}$ is the CPSD matrix of the target signal ($\mathbb{C}^{M \times M}$),
 $\mathbf{P}_{u_i} = \mathbf{E} \{ u_i \mathbf{b}_i \mathbf{b}_i^H u_i^* \}$ is the CPSD matrix of the i^{th} interfering signal ($\mathbb{C}^{M \times M}$),
 $\mathbf{P}_v = \mathbf{E} \{ \mathbf{v}\mathbf{v}^H \}$ is the CPSD matrix of the background noise signal ($\mathbb{C}^{M \times M}$), and
 \mathbf{P} is the CPSD matrix of the noise ($\mathbb{C}^{M \times M}$).

2.2.2 Binaural Cues

To localise sound in the horizontal plane, the binaural cues, more specifically, the ILDs and the ITDs need to be preserved [10], [12]. In the frequency domain, the ITD corresponds to interaural phase difference (IPD). These cues can be obtained from the interaural transfer function (ITF). Let the ITF of the target before and after processing, i.e., at the input and the output of the filter, be defined as

$$\text{ITF}_x^{in} = \frac{a_1}{a_M}, \quad \text{ITF}_x^{out} = \frac{\mathbf{w}_L^H \mathbf{a}}{\mathbf{w}_R^H \mathbf{a}}. \quad (2.7)$$

Similarly, we can define the input and the output ITF of the i^{th} interfering source as

$$\text{ITF}_{u_i}^{in} = \frac{b_{i,1}}{b_{i,M}}, \quad \text{ITF}_{u_i}^{out} = \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i}. \quad (2.8)$$

From the ITF, we can also obtain the ILDs of the target and the interfering sources, before and after processing as

$$\begin{aligned} \text{ILD}_x^{in} &= |\text{ITF}_x^{in}|^2, & \text{ILD}_x^{out} &= |\text{ITF}_x^{out}|^2, \\ \text{ILD}_{u_i}^{in} &= |\text{ITF}_{u_i}^{in}|^2, & \text{ILD}_{u_i}^{out} &= |\text{ITF}_{u_i}^{out}|^2, \end{aligned} \quad (2.9)$$

as well as, the IPDs

$$\begin{aligned} \text{IPD}_x^{in} &= \angle \text{ITF}_x^{in}, & \text{IPD}_x^{out} &= \angle \text{ITF}_x^{out}, \\ \text{IPD}_{u_i}^{in} &= \angle \text{ITF}_{u_i}^{in}, & \text{IPD}_{u_i}^{out} &= \angle \text{ITF}_{u_i}^{out}. \end{aligned} \quad (2.10)$$

In order to preserve the binaural cues, the ILDs and the ITDs at the input and the output must be the same. The ITF error for the target and the interferers can be expressed as

$$\begin{aligned} \gamma_x &= |\text{ITF}_x^{in} - \text{ITF}_x^{out}|, \\ \gamma_{u_i} &= |\text{ITF}_{u_i}^{in} - \text{ITF}_{u_i}^{out}|. \end{aligned} \quad (2.11)$$

While the ILD errors can be gives as

$$\begin{aligned}\epsilon_x &= |\text{ILD}_x^{in} - \text{ILD}_x^{out}|, \\ \epsilon_{u_i} &= |\text{ILD}_{u_i}^{in} - \text{ILD}_{u_i}^{out}|,\end{aligned}\tag{2.12}$$

and the ITD errors can be given as

$$\begin{aligned}\tau_x &= \frac{|\text{IPD}_x^{in} - \text{IPD}_x^{out}|}{\pi}, \\ \tau_{u_i} &= \frac{|\text{IPD}_{u_i}^{in} - \text{IPD}_{u_i}^{out}|}{\pi}.\end{aligned}\tag{2.13}$$

In this work, only spatial filtering algorithms are considered, i.e., the target estimate is constrained to be undistorted. Hence, the ILD and the ITD cues of the target are always preserved [21].

2.2.3 Degrees of freedom (DoF)

The degrees of freedom (DoF) are the maximum number of independent variables available in the model, that must be specified to determine the feasible solution of the problem. In the algorithms discussed in Section 2.3, the DoF determine their noise reduction ability.

The maximum available degrees is $2M$ for an unconstrained noise power minimisation problem. This reduces by one, with every equality constraint introduced. The DoF available for noise reduction in the algorithms discussed, are mentioned in their corresponding sections.

2.3 Previous Work

2.3.1 Binaural minimum variance distortionless response (BMVDR)

The MVDR maximises the signal-to-noise ratio (SNR) in the target direction, while maintaining the target signal undistorted [4]. The BMVDR is a binaural extension of the MVDR beamforming algorithm.

The BMVDR consists of two spatial filters, say, \mathbf{w}_L and \mathbf{w}_R , that are estimated by minimising the output interference and background noise power, while constraining the target to be undistorted. The optimisation problem can written as done in [21],

$$\begin{aligned}\mathbf{w}_{\text{BMVDR},L} &= \arg \min_{\mathbf{w}_L \in \mathbb{C}^{M \times 1}} \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L \\ &\text{subject to } \mathbf{w}_L^H \mathbf{a} = a_1,\end{aligned}\tag{2.14}$$

$$\begin{aligned}\mathbf{w}_{\text{BMVDR},R} &= \arg \min_{\mathbf{w}_R \in \mathbb{C}^{M \times 1}} \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R \\ &\text{subject to } \mathbf{w}_R^H \mathbf{a} = a_M.\end{aligned}$$

The closed form solutions can be given by

$$\mathbf{w}_{\text{BMVDR},L} = \frac{\mathbf{P}^{-1}\mathbf{a}a_1^*}{\mathbf{a}^H\mathbf{P}^{-1}\mathbf{a}}, \quad \mathbf{w}_{\text{BMVDR},R} = \frac{\mathbf{P}^{-1}\mathbf{a}a_M^*}{\mathbf{a}^H\mathbf{P}^{-1}\mathbf{a}}. \quad (2.15)$$

The two constrained problems in Eq. (2.14) can also be combined, and jointly written as

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\text{minimise}} && \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{P} & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{P} \end{bmatrix}}_{\tilde{\mathbf{P}} \in \mathbb{C}^{2M \times 2M}} \mathbf{w} \\ & \text{subject to} && \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix}}_{\Lambda_A \in \mathbb{C}^{2M \times 2}} = \underbrace{\begin{bmatrix} a_1 & a_M \end{bmatrix}}_{\mathbf{f}_A^H \in \mathbb{C}^{1 \times 2}}. \end{aligned} \quad (2.16)$$

The closed form solution to Eq. (2.16) is

$$\mathbf{w}_{\text{BMVDR}} = \tilde{\mathbf{P}}^{-1} \Lambda_A \left(\Lambda_A^H \tilde{\mathbf{P}}^{-1} \Lambda_A \right)^{-1} \mathbf{f}_A, \quad (2.17)$$

where $\mathbf{w}_{\text{BMVDR}} = \left[\mathbf{w}_{\text{BMVDR},L}^H \quad \mathbf{w}_{\text{BMVDR},R}^H \right]^H \in \mathbb{C}^{2M \times 1}$.

Since the target signal is undistorted due to the distortionless constraints, the spatial cues of the target are preserved. However, the spatial cues of the interferers become identical to those of the target [21]. Hence the interferers appear to be co-located with the target after processing. This can be shown by calculating the output ITF of the target and the interfering sources, which become identical to the ITF of the target at the input of the beam-former. The ITF of the target at the output can be written as

$$\text{ITF}_x^{out} = \frac{\mathbf{w}_{\text{BMVDR},L}^H \mathbf{a}}{\mathbf{w}_{\text{BMVDR},R}^H \mathbf{a}} = \frac{a_1}{a_M} = \text{ITF}_x^{in}, \quad (2.18)$$

while for the interferers

$$\text{ITF}_{n_i}^{out} = \frac{\mathbf{w}_{\text{BMVDR},L}^H \mathbf{b}_i}{\mathbf{w}_{\text{BMVDR},R}^H \mathbf{b}_i} = \frac{a_1}{a_M} = \text{ITF}_x^{in}. \quad (2.19)$$

With the BMVDR, there are only two constraints. Hence the DoF available for noise reduction are $D = 2M - 2$. This allows for maximum noise reduction, and hence maximum SNR, in comparison to the remaining methods discussed.

Although the SNR is maximised, the spatial impression of the interferers and the background noise are not maintained with the BMVDR algorithm. Hence additional constraints, that preserve the spatial cues of the interferers, can be used as explained in the following algorithms.

2.3.2 Binaural linearly constrained minimum variance (BLCMV)

Since the BMVDR fails to preserve the spatial cues of the interferers, in [22] additional linear constraints were introduced to help to preserve the spatial cues of the interferers. In the proposed BLCMV, the spatial cues of the interferers are preserved by suppressing the interferers using a pre-determined rejection parameter. This is done by introducing the following constraints to the BMVDR problem in Eq. (2.14).

$$\mathbf{w}_L^H \mathbf{b}_i = \eta_L b_{i,1}, \quad \mathbf{w}_R^H \mathbf{b}_i = \eta_R b_{i,M}, \quad i = 1, 2, \dots, r, \quad (2.20)$$

where η_L and $\eta_R \in \mathcal{R}$ are the rejection parameters at the left and right ears respectively.

The BLCMV joint optimisation problem with the constraints in Eq. (2.20) can be written as

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\text{minimise}} \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{P} & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{P} \end{bmatrix}}_{\tilde{\mathbf{P}} \in \mathbb{C}^{2M \times 2M}} \mathbf{w} \\ & \text{subject to } \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{\Lambda}_A & \mathbf{\Lambda}_B \end{bmatrix}}_{\mathbf{\Lambda} \in \mathbb{C}^{2M \times 2r+2}} = \underbrace{\begin{bmatrix} \mathbf{f}_A^H & \mathbf{f}_B^H \end{bmatrix}}_{\mathbf{f}^H \in \mathbb{C}^{1 \times 2r+2}} \\ & \text{where } \mathbf{\Lambda}_A = \begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix} \in \mathbb{C}^{2M \times 2}, \\ & \mathbf{\Lambda}_B = \begin{bmatrix} \mathbf{b}_1 & \mathbf{0} & \dots & \mathbf{b}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{b}_1 & \dots & \mathbf{0} & \mathbf{b}_r \end{bmatrix} \in \mathbb{C}^{2M \times 2r}, \\ & \mathbf{f}_A^H = [a_1 \quad a_M] \in \mathbb{C}^{1 \times 2}, \\ & \mathbf{f}_B^H = [\eta_L b_{1,1} \quad \eta_R b_{1,M} \quad \dots \quad \eta_L b_{r,1} \quad \eta_R b_{r,M}] \in \mathbb{C}^{1 \times 2r}. \end{aligned} \quad (2.21)$$

The closed form solution to Eq. (2.21) is given by

$$\mathbf{w}_{\text{BLCMV}} = \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda}^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \right)^{-1} \mathbf{f}, \quad \text{for } r \leq r_{\max}. \quad (2.22)$$

Here, the ITF of the target and the interferers at the output are given by

$$\text{ITF}_x^{\text{out}} = \frac{\mathbf{w}_{\text{BLCMV},L}^H \mathbf{a}}{\mathbf{w}_{\text{BLCMV},R}^H \mathbf{a}} = \frac{a_1}{a_M} = \text{ITF}_x^{\text{in}}, \quad (2.23)$$

$$\text{ITF}_{u_i}^{\text{out}} = \frac{\mathbf{w}_{\text{BLCMV},L}^H \mathbf{b}_i}{\mathbf{w}_{\text{BLCMV},R}^H \mathbf{b}_i} = \frac{\eta_L b_{i,1}}{\eta_R b_{i,M}} = \frac{\eta_L}{\eta_R} \text{ITF}_{u_i}^{\text{in}}.$$

By taking $\eta_L = \eta_R = \eta$, the ITF of the interferers can be preserved. An optimum choice of η is proposed in [27].

With the BLCMV, there are two additional constraints per interferer introduced, in addition to the two target distortionless constraints, as used with the BMVDR. Hence, for ‘ r ’ interferers, the DoF available for noise reduction are $D = 2M - 2r - 2$. It is required that at least one DoF is available for noise reduction, if not, the objective of the problem cannot be optimised. Therefore, the BLCMV can preserve the spatial cues of maximum $r_{\max} = M - 2$ interferers.

2.3.3 Joint binaural linearly constrained minimum variance (JBLCMV)

With the BLCMV, two constraints were per interferer were introduced to preserve their ILDs and ITDs. However, the maximum number of interferers r_{max} is limited by the number of microphones used in the hearing aids. Commonly, HAs use up to $\frac{M}{2} = 3$ microphones per unit, that limits r_{max} to 4 interferers. In order to tackle this drawback, a modified constraint can be introduced as proposed in [23].

Here, the principle of ITF preservation, as mentioned in Eq. (2.8), is used to derive a linear constraint per interferer, that is

$$\text{ITF}_{u_i}^{out} = \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} = \frac{b_{i,1}}{b_{i,M}} = \text{ITF}_{u_i}^{in}, \quad i = 1, \dots, r, \quad (2.24)$$

which can be re-written as,

$$\mathbf{w}_L^H \mathbf{b}_i b_{i,M} - \mathbf{w}_R^H \mathbf{b}_i b_{i,1} = 0 \quad i = 1, \dots, r. \quad (2.25)$$

Using the linear constraints in Eq. (2.25) for each interferer, in combination with the BMVDR formulation, gives the following joint optimisation problem.

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\text{minimise}} \quad \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\ & \text{subject to} \quad \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{\Lambda}_A & \mathbf{\Lambda}_C \end{bmatrix}}_{\mathbf{\Lambda}} = \underbrace{\begin{bmatrix} \mathbf{f}_A^H & \mathbf{f}_C^H \end{bmatrix}}_{\mathbf{f}^H}, \\ & \text{where} \quad \mathbf{\Lambda}_C = \begin{bmatrix} \mathbf{b}_1 b_{1,M} & \dots & \mathbf{b}_r b_{r,M} \\ -\mathbf{b}_1 b_{1,1} & \dots & -\mathbf{b}_r b_{r,1} \end{bmatrix} \in \mathbb{C}^{2M \times r}, \\ & \quad \mathbf{f}_C^H = [0 \quad \dots \quad 0] \in \mathbb{C}^{1 \times r}. \end{aligned} \quad (2.26)$$

The closed form solution to Eq. (2.26) is given by

$$\mathbf{w}_{\text{JBLCMV}} = \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \left(\mathbf{\Lambda}^H \tilde{\mathbf{P}}^{-1} \mathbf{\Lambda} \right)^{-1} \mathbf{f}, \quad \text{for } r \leq r_{max}. \quad (2.27)$$

With the JBLCMV, there is one additional constraint introduced per interferer, in addition to the two target distortionless constraints of the BMVDR. Hence for ‘ r ’ interferers, the DoF available for noise reduction are $D = 2M - r - 2$. The maximum interferers for which the spatial cues can be preserved is $r_{max} = 2M - 3$. Hence, more interferers can be considered with the JBLCMV in comparison to the BLCMV. Moreover, the noise reduction is better with the JBLCMV than the BLCMV for the same number of interferers.

2.3.4 Relaxed binaural linearly constrained minimum variance (RBLCMV)

With the BLCMV and the JBLCMV, the ITF cues are preserved perfectly, however, they are limited by the number of interferers that can be constrained. In [26], the

authors proposed a binaural beam-former that allows for a bounded ITF error of the interferers, that is,

$$\begin{aligned}
& \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\text{minimise}} \quad \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\
& \text{subject to} \quad \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\
& \quad \underbrace{\left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} - \frac{b_{i,1}}{b_{i,M}} \right|}_{\gamma_{\mathbf{u}_i}} \leq e_i, \quad i = 1, \dots, r, \\
& \quad \text{where} \quad e_i = c_i \underbrace{\left| \frac{a_1}{a_M} - \frac{b_{i,1}}{b_{i,M}} \right|}_{\gamma_{\mathbf{u}_i}^{\text{BMVDR}}}, \\
& \quad \text{and} \quad 0 \leq c_i \leq 1.
\end{aligned} \tag{2.28}$$

Here, the ITF errors of the interferers are upper bound by a factor of the ITF errors observed with the BMVDR. Since Eq. (2.28) is not a convex formulation, no closed form expression can be derived. The problem was solved approximately, using successive convex optimisation. By replacing the equality constraints in Eq. (2.26) with inequality constraints, the feasibility set widens, and allows the spatial cue preservation for more interferers. Moreover, for a given number of interferers, the DoF increases, improving the noise reduction performance.

Since the problem has inequality constraints, the DoF can not be calculated in this case.

With the RBLCMV beam-former, the noise reduction performance improves due to bounded errors in the binaural cues. Drawing motivation from it, and using the principle of the Duplex Theory, Chapter 3 aims to formulate an optimisation problem, that preserves the ILD cues of the noise component, while not controlling the ITD cues of the noise component.

Methods Proposed

This chapter presents the two methods proposed to perform binaural beamforming with interaural level difference (ILD) cue preservation. The problem formulation and the convexity of the two methods are discussed. Additionally, the steps taken to find a convex relaxation that solves the problem in polynomial time are discussed.

3.1 Interaural Level Difference Preservation Methods

In the following sections, the cue preservation methods use constraints, in addition to the original formulation of the binaural minimum variance distortionless response (BMVDR) beam-former, that preserve only the ILD cues after processing.

3.1.1 Perfect interaural level difference cue preservation (P-ILD)

In this formulation of the optimisation problem, the aim is to perfectly preserve the ILD cues of the interferers. In addition to the target distortionless constraints, one constraint per interferer is introduced to preserve the ILD cues. The problem formulation is given by

$$\begin{aligned}
 & \underset{\mathbf{w}_L, \mathbf{w}_R \in \mathbb{C}^{M \times 1}}{\text{minimise}} && \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L + \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R \\
 & \text{subject to} && \mathbf{w}_L^H \mathbf{a} = a_1 \quad \mathbf{w}_R^H \mathbf{a} = a_M \\
 & && \underbrace{\left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} \right|^2}_{\text{ILD}_{u_i}^{\text{out}}} - \underbrace{\left| \frac{b_{i,1}}{b_{i,M}} \right|^2}_{\text{ILD}_{u_i}^{\text{in}}} = 0, \quad i = 1, \dots, r \leq r_{\max}.
 \end{aligned} \tag{3.1}$$

In Eq. (3.1), the objective function and the first two equality constraints follow the principle of the BMVDR, i.e., the objective function minimises the output noise power, while the two equality constraints keep the target signal undistorted in both ears. The additional equality constraints keep the $\text{ILD}_{u_i}^{\text{out}}$, the output ILD of the i^{th} interferer, equal to the $\text{ILD}_{u_i}^{\text{in}}$, the input ILD of the i^{th} interferer.

By expanding the ILD equality constraints, the problem can be jointly optimised with respect to \mathbf{w}_L and \mathbf{w}_R .

That is, starting from

$$\frac{\mathbf{w}_L^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_L}{\mathbf{w}_R^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_R} - \frac{|b_{i,1}|^2}{|b_{i,M}|^2} = 0, \quad i = 1, \dots, r \leq r_{\max},$$

and by cross multiplying, the constraint can be re-written as

$$\mathbf{w}_L^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_L |b_{i,M}|^2 - \mathbf{w}_R^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_R |b_{i,1}|^2 = 0.$$

The ILD constraint can now be written into a matrix equation as

$$\underbrace{\begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix}}_{\mathbf{w}^H} \underbrace{\begin{bmatrix} \mathbf{b}_i \mathbf{b}_i^H |b_{i,M}|^2 & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & -\mathbf{b}_i \mathbf{b}_i^H |b_{i,1}|^2 \end{bmatrix}}_{\mathbf{M}_i \in \mathbb{C}^{2M \times 2M}} \underbrace{\begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix}}_{\mathbf{w}} = 0. \quad (3.2)$$

Using Eq. (3.2) to replace the ILD constraints in Eq. (3.1), the problem can be re-written as

$$\begin{aligned} (\mathbf{P}_1) \quad & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\text{minimise}} \quad \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{P} & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{P} \end{bmatrix}}_{\tilde{\mathbf{P}} \in \mathbb{C}^{2M \times 2M}} \mathbf{w} \\ & \text{subject to} \quad \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{a} & \mathbf{0} \\ \mathbf{0} & \mathbf{a} \end{bmatrix}}_{\Lambda_A \in \mathbb{C}^{2M \times 2}} = \underbrace{\begin{bmatrix} a_1 & a_M \end{bmatrix}}_{\mathbf{f}_A^H \in \mathbb{C}^{1 \times 2}} \\ & \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{b}_i \mathbf{b}_i^H |b_{i,M}|^2 & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & -\mathbf{b}_i \mathbf{b}_i^H |b_{i,1}|^2 \end{bmatrix}}_{\mathbf{M}_i \in \mathbb{C}^{2M \times 2M}} \mathbf{w} = 0, \quad i = 1, \dots, r \leq r_{max}. \end{aligned} \quad (3.3)$$

Throughout this chapter, let \mathbf{p}_1^* denote the optimal value of the objective function in Eq. (3.3), and let \mathbf{w}_1^* refer to the optimal point, i.e., the point at which the optimal value \mathbf{p}_1^* is attained.

Eq. (3.3) is a quadratically constrained quadratic program (QCQP) problem, since the objective function is quadratic in \mathbf{w} , with linear and quadratic equality constraints. The objective function is convex since $\tilde{\mathbf{P}}$ is positive semi-definite (PSD) (A.2). The problem, however, is a non-convex optimisation problem due to the quadratic equality constraints on the interferers.

Non-convexity of the Eq. (3.3):

- Objective : Quadratic in ‘ \mathbf{w} ’
- Target Equality Constraint : Linear in ‘ \mathbf{w} ’
- **Interferer Equality Constraints** : Quadratic in ‘ \mathbf{w} ’

In general, QCQPs are known to be NP-hard [28], except for certain special cases which show hidden convexity [29],[30],[31]. Here, Eq. (3.3) is a non-convex QCQP and it is hard to compute a global optimal solution due to its NP-hardness.

Such non-convex QCQP problems, are commonly overcome by implementing efficient approximation techniques using semi-definite relaxations as explained in the following section [32].

3.1.1.1 Convex Relaxation

Consider a matrix $\mathbf{W} = \mathbf{w}\mathbf{w}^H$. To linearise the quadratic equality constraint, the problem in Eq. (3.3) can be re-written in terms of \mathbf{W} (A.1).

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}, \mathbf{W} \in \mathbb{C}^{2M \times 2M}}{\text{minimise}} && \text{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\ & \text{subject to} && \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\ & && \text{Tr}(\mathbf{W}\mathbf{M}_i) = 0, \quad i = 1, \dots, r \leq r_{max}, \\ & && \mathbf{W} = \mathbf{w}\mathbf{w}^H. \end{aligned} \quad (3.4)$$

In Eq. (3.4), the non-convex quadratic equality constraint of Eq. (3.3) is linearised. The problem, however, remains non-convex due to the newly introduced quadratic equality constraint, $\mathbf{W} = \mathbf{w}\mathbf{w}^H$.

Commonly, the prevailing method used to overcome this is by relaxing the equivalence constraint $\mathbf{W} = \mathbf{w}\mathbf{w}^H$ to $\mathbf{W} \succeq \mathbf{w}\mathbf{w}^H$. By relaxing the constraint, the problem is now convex as shown in Eq. (3.5).

$$\begin{aligned} & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}, \mathbf{W} \in \mathbb{C}^{2M \times 2M}}{\text{minimise}} && \text{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\ & \text{subject to} && \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\ & && \text{Tr}(\mathbf{W}\mathbf{M}_i) = 0, \quad i = 1, \dots, r \leq r_{max}, \\ & && \mathbf{W} \succeq \mathbf{w}\mathbf{w}^H. \end{aligned} \quad (3.5)$$

Since $\mathbf{W} \succeq \mathbf{w}\mathbf{w}^H$ can be written as a linear matrix inequality $\mathbf{W} - \mathbf{w}\mathbf{w}^H \succeq 0$, the problem can be re-written as a semi-definite program (SDP). Since the non-convex problem in Eq. (3.4) can be relaxed into a convex SDP, it is referred to as a semi-definite relaxation (SDR). That is,

$$\begin{aligned} (\text{SDR}_1) \quad & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}, \mathbf{W} \in \mathbb{C}^{2M \times 2M}}{\text{minimise}} && \text{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\ & \text{subject to} && \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\ & && \text{Tr}(\mathbf{W}\mathbf{M}_i) = 0, \quad i = 1, \dots, r \leq r_{max}, \\ & && \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & 1 \end{bmatrix} \succeq 0. \end{aligned} \quad (3.6)$$

By replacing the original set of constraints in Eq. (3.4), with the new set of constraints in Eq. (3.6), the feasibility set is broadened. Having a broader feasibility region, the optimal value $\mathbf{p}_{\text{SDR}_1}^*$, of Eq. (3.6) may be lower or equal to the optimal value \mathbf{p}^* of the original problem in Eq. (3.4). That is, $\mathbf{p}_{\text{SDR}_1}^*$ gives a lower bound to the optimal value \mathbf{p}_1^* of the original problem [33]. Moreover, it can be proved that $\mathbf{p}_{\text{SDR}_1}^*$ will also be the optimal value to the dual of \mathbf{P}_1 (A.3), hence providing a non-trivial lower bound [34], that is

$$\mathbf{p}_1^* \geq \mathbf{p}_{\text{SDR}_1}^* \quad (3.7)$$

Let us consider $(\mathbf{W}_{\text{SDR}_1}^*, \mathbf{w}_{\text{SDR}_1}^*)$ to be the optimal points of Eq. (3.6). The relaxed problem SDR_1 , will be equivalent to the original problem \mathbf{P}_1 , only when $\mathbf{W}_{\text{SDR}_1}^* = \mathbf{w}_{\text{SDR}_1}^* \mathbf{w}_{\text{SDR}_1}^{*H}$. If this happens, the optimal solution of the problem SDR_1 , is the global optimal solution of the non-convex QCQP in Eq. (3.3). However if this equivalence is not met, then the optimal point $(\mathbf{W}_{\text{SDR}_1}^*, \mathbf{w}_{\text{SDR}_1}^*)$ is not a feasible point to the problem in Eq. (3.4).

Nevertheless, as mentioned earlier, the optimal value $\mathbf{p}_{\text{SDR}_1}^*$ is non-trivial. It provides an over-estimation of the noise reduction ability, which can be used to analyse the performance of the method proposed, as done in Chapter 4. Furthermore, this lower bound can be tightened by using additional redundant constraints that satisfy the problem \mathbf{P}_1 . One way to perform this, is using the reformulation-linearisation technique (RLT) proposed in [35]. Here, the product of the linear constraints in Eq. (3.1) are linearised as follows.

Pre-multiplying \mathbf{w} with the target distortionless constraint we get

$$\mathbf{w} \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{w} \mathbf{f}_A^H,$$

and by linearising with \mathbf{W}

$$\mathbf{W} \boldsymbol{\Lambda}_A - \mathbf{w} \mathbf{f}_A^H = 0. \quad (3.8)$$

Multiplying the target distortionless constraint with itself we get

$$(\mathbf{w}^H \boldsymbol{\Lambda}_A - \mathbf{f}_A^H) (\mathbf{w}^H \boldsymbol{\Lambda}_A - \mathbf{f}_A^H)^H = 0,$$

on expanding which

$$\mathbf{w}^H (\boldsymbol{\Lambda}_A \boldsymbol{\Lambda}_A^H) \mathbf{w} - \mathbf{w}^H \boldsymbol{\Lambda}_A \mathbf{f}_A - \mathbf{f}_A^H \boldsymbol{\Lambda}_A^H \mathbf{w} + \mathbf{f}_A^H \mathbf{f}_A = 0,$$

and by linearising with \mathbf{W}

$$\text{Tr}(\mathbf{W} \boldsymbol{\Lambda}_A \boldsymbol{\Lambda}_A^H) - \mathbf{w}^H \boldsymbol{\Lambda}_A \mathbf{f}_A - (\boldsymbol{\Lambda}_A \mathbf{f}_A)^H \mathbf{w} + \mathbf{f}_A^H \mathbf{f}_A = 0. \quad (3.9)$$

The constraints in Eq. (3.8) and Eq. (3.9) are redundant, as they are formulated from the original distortionless target constraint. By introducing these constraints, the lower bound of the problem \mathbf{P}_1 , provided by the optimal value $\mathbf{p}_{\text{SDR}_1}^*$ of SDR_1 , is tightened further and the approximation is improved.

$$\begin{aligned} (\text{SDR-RLT}_1) \quad & \underset{\substack{\mathbf{W} \in \mathbb{C}^{2M \times 2M}, \\ \mathbf{w} \in \mathbb{C}^{2M \times 1}}}{\text{minimise}} \quad \text{Tr}(\mathbf{W} \tilde{\mathbf{P}}) \\ & \text{subject to } \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\ & \quad \text{Tr}(\mathbf{W} \mathbf{M}_i) = 0, \quad i = 1, \dots, r \leq r_{max}, \\ & \quad \mathbf{W} \boldsymbol{\Lambda}_A - \mathbf{w} \mathbf{f}_A^H = 0 \\ & \quad \text{Tr}(\mathbf{W} \boldsymbol{\Lambda}_A \boldsymbol{\Lambda}_A^H) - \mathbf{w}^H \boldsymbol{\Lambda}_A \mathbf{f}_A - (\boldsymbol{\Lambda}_A \mathbf{f}_A)^H \mathbf{w} + \mathbf{f}_A^H \mathbf{f}_A = 0 \\ & \quad \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & 1 \end{bmatrix} \succeq 0. \end{aligned} \quad (3.10)$$

Taking the optimal value of the problem $\text{SDR} - \text{RLT}_1$ as $\mathbf{p}_{\text{SDR-RLT}_1}^*$, the relationship between the optimal solutions among the relaxations can be given as

$$\mathbf{p}_1^* \geq \mathbf{p}_{\text{SDR-RLT}_1}^* \geq \mathbf{p}_{\text{SDR}_1}^*. \quad (3.11)$$

3.1.1.2 Degrees of freedom (DoF)

In problem \mathbf{P}_1 , there is one constraint per interferer introduced, in addition to the two target distortionless constraints of the BMVDR. Hence for ‘ r ’ interferers the DoF available for noise reduction are $D = 2M - r - 2$. Assuming that there is one DoF for noise reduction, the maximum number of interferers r_{max} can be $2M - 3$.

However since the problem $\mathbf{SDR} - \mathbf{RLT}_1$ is used to find an approximate solution to the non-convex problem \mathbf{P}_1 , the performance is expected to deviate. With every interferer cue preservation constraint, the ILD errors are expected to get worse. This is because, the optimality gap between \mathbf{P}_1 and $\mathbf{SDR} - \mathbf{RLT}_1$ increases and thereby worsening the constraint violations by the solution $\mathbf{w}^*_{\mathbf{SDR}-\mathbf{RLT}_1}$.

3.1.2 Relaxed interaural level difference cue preservation (R-ILD)

In this formulation, instead of preserving the ILD cues perfectly, the cue errors are upper bound by a suitable factor.

$$\begin{aligned} & \underset{\mathbf{w}_L, \mathbf{w}_R \in \mathbb{C}^{M \times 1}}{\text{minimise}} && \mathbf{w}_L^H \mathbf{P} \mathbf{w}_L + \mathbf{w}_R^H \mathbf{P} \mathbf{w}_R \\ & \text{subject to} && \mathbf{w}_L^H \mathbf{a} = a_1 \quad \mathbf{w}_R^H \mathbf{a} = a_M \\ & && \left| \frac{\mathbf{w}_L^H \mathbf{b}_i}{\mathbf{w}_R^H \mathbf{b}_i} \right|^2 - \left| \frac{b_{i,1}}{b_{i,M}} \right|^2 \leq \mathcal{E}_i, \quad i = 1, \dots, r \leq r_{max}. \end{aligned} \quad (3.12)$$

The parameter \mathcal{E}_i can be chosen suitably based on the acceptable ILD error. Here, \mathcal{E}_i is chosen in a way that the ILD cue errors are lower than that found with the BMVDR algorithm, that is,

$$\mathcal{E}_i = c_i \underbrace{\left| \frac{a_1}{a_M} \right|^2 - \left| \frac{b_{i,1}}{b_{i,M}} \right|^2}_{\epsilon_{u_i}^{\text{BMVDR}}}, \quad (3.13)$$

$$\text{with } 0 \leq c_i \leq 1.$$

The choice of ‘ c_i ’ is made based on the best trade-off between ILD cue preservation and noise reduction achieved.

The ILD inequality constraint can be expanded as follows,

$$\begin{aligned} & \frac{|\mathbf{w}_L^H \mathbf{b}_i|^2}{|\mathbf{w}_R^H \mathbf{b}_i|^2} - \frac{|b_{i,1}|^2}{|b_{i,M}|^2} \leq \mathcal{E}_i, \quad i = 1, \dots, r \leq r_{max}, \\ & \text{and} \\ & - \left(\frac{|\mathbf{w}_L^H \mathbf{b}_i|^2}{|\mathbf{w}_R^H \mathbf{b}_i|^2} - \frac{|b_{i,1}|^2}{|b_{i,M}|^2} \right) \leq \mathcal{E}_i, \quad i = 1, \dots, r \leq r_{max}. \end{aligned} \quad (3.14)$$

This can be expanded as

$$\begin{aligned} & \mathbf{w}_L^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_L |b_{i,M}|^2 - \mathbf{w}_R^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_R (|b_{i,1}|^2 + \mathcal{E}_i |b_{i,M}|^2) \leq 0, \\ \text{and,} & \\ & -\mathbf{w}_L^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_L |b_{i,M}|^2 + \mathbf{w}_R^H \mathbf{b}_i \mathbf{b}_i^H \mathbf{w}_R (|b_{i,1}|^2 - \mathcal{E}_i |b_{i,M}|^2) \leq 0. \end{aligned} \quad (3.15)$$

By using the constraints in Eq. (3.15) to replace the ILD inequality constraints in Eq. (3.12) and jointly optimising with respect to \mathbf{w}_L and \mathbf{w}_R , it is once again a non-convex QCQP, similar to the previous method P-ILD.

$$\begin{aligned} (\mathbf{P}_2) \quad & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\text{minimise}} \quad \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\ & \text{subject to} \quad \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\ & \mathbf{w}^H \underbrace{\begin{bmatrix} \mathbf{b}_i \mathbf{b}_i^H |b_{i,M}|^2 & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & -\mathbf{b}_i \mathbf{b}_i^H (|b_{i,1}|^2 + \mathcal{E}_i |b_{i,M}|^2) \end{bmatrix}}_{\mathbf{M}_{A,i} \in \mathbb{C}^{2M \times 2M}} \mathbf{w} \leq 0, \quad i = 1, \dots, r \leq r_{max}, \\ & \mathbf{w}^H \underbrace{\begin{bmatrix} -\mathbf{b}_i \mathbf{b}_i^H |b_{i,M}|^2 & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{b}_i \mathbf{b}_i^H (|b_{i,1}|^2 - \mathcal{E}_i |b_{i,M}|^2) \end{bmatrix}}_{\mathbf{M}_{B,i} \in \mathbb{C}^{2M \times 2M}} \mathbf{w} \leq 0, \quad i = 1, \dots, r \leq r_{max}. \end{aligned} \quad (3.16)$$

Once again, the problem \mathbf{P}_2 is a non-convex QCQP. Similar to problem \mathbf{P}_1 , an approximate solution can be found using a SDR with RLT.

Non-convexity of the Eq. (3.14):

- Objective : Quadratic in ‘ \mathbf{w} ’
- Target Equality Constraint : Linear in ‘ \mathbf{w} ’
- **Interferer In-equality Constraints** : Non-convex quadratic functions in ‘ \mathbf{w} ’ since $\mathbf{M}_{A,i}$ & $\mathbf{M}_{B,i}$ are not PSD

3.1.2.1 Convex Relaxation

The convex relaxation of the problem \mathbf{P}_2 , using linear matrix inequalities and RLT, as done with \mathbf{P}_1 can be given as

$$\begin{aligned} (\text{SDR-RLT}_2) \quad & \underset{\substack{\mathbf{W} \in \mathbb{C}^{2M \times 2M}, \\ \mathbf{w} \in \mathbb{C}^{2M \times 1}}}{\text{minimise}} \quad \text{Tr}(\mathbf{W} \tilde{\mathbf{P}}) \\ & \text{subject to} \quad \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\ & \quad \text{Tr}(\mathbf{W} \mathbf{M}_{A,i}) \leq 0, \quad i = 1, \dots, r \leq r_{max}, \\ & \quad \text{Tr}(\mathbf{W} \mathbf{M}_{B,i}) \leq 0, \quad i = 1, \dots, r \leq r_{max}, \\ & \quad \mathbf{W} \boldsymbol{\Lambda}_A - \mathbf{w} \mathbf{f}_A^H = 0 \\ & \quad \text{Tr}(\mathbf{W} \boldsymbol{\Lambda}_A \boldsymbol{\Lambda}_A^H) - \mathbf{w}^H \boldsymbol{\Lambda}_A \mathbf{f}_A - (\boldsymbol{\Lambda}_A \mathbf{f}_A)^H \mathbf{w} + \mathbf{f}_A^H \mathbf{f}_A = 0 \\ & \quad \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & 1 \end{bmatrix} \succeq 0. \end{aligned} \quad (3.17)$$

3.1.2.2 Degrees of freedom (DoF)

In problem \mathbf{P}_2 , the feasible set is larger than in problem \mathbf{P}_1 and a limit in terms of r_{max} cannot be predicted.

Similar to problem $\mathbf{SDR} - \mathbf{RLT}_1$, by using the approximate solution of the problem $\mathbf{SDR} - \mathbf{RLT}_2$ to problem \mathbf{P}_2 , the constraint violations of the solution $\mathbf{w}^*_{\mathbf{SDR}-\mathbf{RLT}_2}$ with respect to the ILD preservation gets worse with increasing ‘ r ’.

As mentioned previously, problems \mathbf{P}_1 , \mathbf{P}_2 are both non-convex QCQPs. Therefore only approximate solutions are found for both methods proposed and they are not globally optimal for the original problem. However on running simulations it is found that the approximation nearly meets the feasibility constraints of \mathbf{P}_1 , \mathbf{P}_2 , implying that it is a good approximation. In Chapter 4, the optimisation problems discussed here are solved in polynomial time, using the CVX toolbox in Matlab [36].

In this chapter, the methods proposed in Chapter 3 are experimentally evaluated. Section 4.1 describes the acoustic setup and the synthesis of the signals prior to running simulations. The metrics used to evaluate the performance of the methods proposed, are explained in Section 4.2. In Section 4.3, the simulation results are reported, and the performance of sound source localisation against noise reduction is analysed. Lastly, the result of the informal listening test conducted is discussed in Section 4.4.

4.1 Experimental Setup

To analyse the performance of the ILD cue preservation and noise reduction, the simulations compare the following methods

Reference Methods:

1. Binaural minimum variance distortionless response (BMVDR) beam-former
2. Joint binaural linearly constrained minimum variance (JBLCMV) beam-former

Proposed Methods:

1. **Method 1:** JBLCMV ($f < 1.5$ kHz) + proposed P-ILD ($f \geq 1.5$ kHz)
2. **Method 2:** JBLCMV ($f < 1.5$ kHz) + proposed R-ILD ($f \geq 1.5$ kHz)

4.1.1 Acoustic Environment

To simulate a noisy environment with a single desired speaker, a target speech signal \mathbf{s} and ‘ i ’ point interfering speech signals \mathbf{u}_i are taken as shown in Figure 4.1. The sources are placed at $(h, \theta, 0^\circ)$, i.e., at a distance of h , θ azimuthal angle, and 0° elevation from the listener at $(0, 0^\circ, 0^\circ)$. The position of each source is reported in Table 4.1. In the experiments, the interfering signals are considered incrementally, i.e., for 1 interfering source, \mathbf{u}_1 is considered, for 2 interfering sources \mathbf{u}_1 and \mathbf{u}_2 are considered and so on.

To spatialise the point sources reaching the microphones of the hearing aids worn by the listener, the head related impulse responses (HRIRs) from a multi-channel, behind-the-ear (BTE) database are used. For each HA, $\frac{M}{2} = 2$ microphones (the middle and the rear from the database) are considered and the signals are simulated using the BTE HRIR from [37].

The experiments are performed for

1. an anechoic environment with $h = 0.80$ m, and
2. a reverberant office environment with $h = 1$ m.

Table 4.1: Position of the Sources.

Source	Azimuth angle
s	0°
u₁	90°
u₂	-90°
u₃	15°
u₄	-15°
u₅	45°
u₆	-45°
u₇	75°

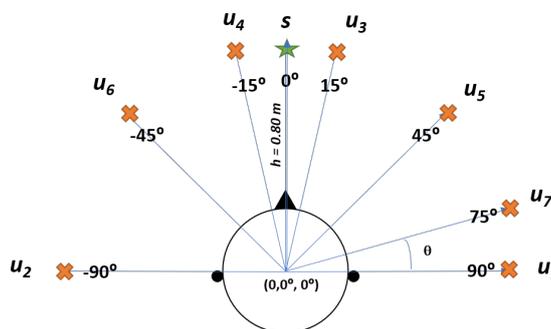


Figure 4.1: Acoustic Environmental Setup.

4.1.2 Signal Synthesis

The speech signal for the target and the interferers are taken from the TIMIT database [38], which is a collection of speech signals in English. The signals are made to be uniformly 30 seconds of duration. The signals are sampled at $f_s = 16$ kHz.

To simulate the setup described in Section 4.1.1, the source signals are convolved with the truncated BTE HRIRs from [37]. For the anechoic scene, the truncated HRIR duration is 50 ms and a duration of 10 ms is used for the office scene. Additionally, to simulate the microphone self noise, white gaussian noise (WGN) at an SNR = 50 dB with respect to the target signal is added to each microphone. Each interfering source is scaled to be 0 dB with respect to the target signal. The signal model in Eq. (2.1) can be written in the discrete time domain as

$$y_m[n] = s[n] * h_{s,m}[n] + \sum_{i=1}^r u_i[n] * h_{i,m}[n] + v_m[n], \quad m = 1, \dots, 4, \quad (4.1)$$

where

$h_{s,m}$ is the impulse response of the target signal from its location to the m^{th} microphone, and

$h_{i,m}$ is the impulse response of the i^{th} interfering signal from its location to the m^{th} microphone.

To perform beam-forming, the signals are transformed to the frequency domain. A short-time Fourier transform (STFT) is applied, taking a window length of 30 ms, applying a square-root Hann window with 50% overlap and 1024 frequency bins. The filtering is done in each frequency bin, per time frame. After filtering, the signals in time domain are synthesised by applying the inverse short-time Fourier transform (ISTFT), followed by applying a square-root Hann window and performing overlap-add synthesis with 50% overlap.

To avoid additional sources of error, such as steering vector mismatch errors, the methods use true ATFs. They are computed by applying the fast Fourier transform (FFT) to the impulse response (IR) with 1024 frequency bins.

4.2 Performance Measures

In this section, the instrumental performance measures that are used to evaluate the methods proposed are described. Section 4.2.1 and Section 4.2.2 present the measures used to assess the behaviour in terms of noise reduction and speech intelligibility, respectively. While in Section 4.2.3, the measures used to assess the preservation of the binaural cues are presented.

4.2.1 Noise Reduction Measures

4.2.1.1 Global segmental signal-to-noise ratio (gsSNR)

Signal-to-noise ratio (SNR) is a primary performance measure, commonly used to assess the quality of speech processing algorithms. It measures the noise reduction performance, by evaluating the ratio of the power of the desired source to the power of the total interfering and noise sources. Here, we use binaural global segmental signal-to-noise ratio (gsSNR), where the entire signal is split into blocks and the binaural SNR is evaluated per block and finally, averaged over the blocks [21],[39].

In the spectro-temporal domain, the binaural SNR in each time-frame ‘ l ’ (block) can be given as

$$\text{SNR}^{\text{in}}(l) = \frac{\sum_{k=1}^N \mathbf{e}^T \tilde{\mathbf{P}}_{\mathbf{x}}(l, k) \mathbf{e}}{\sum_{k=1}^N \mathbf{e}^T \tilde{\mathbf{P}}(l, k) \mathbf{e}}, \quad (4.2)$$

$$\text{SNR}^{\text{out}}(l) = \frac{\sum_{k=1}^N \mathbf{w}^H(l, k) \tilde{\mathbf{P}}_{\mathbf{x}}(l, k) \mathbf{w}(l, k)}{\sum_{k=1}^N \mathbf{w}^H(l, k) \tilde{\mathbf{P}}(l, k) \mathbf{w}(l, k)}, \quad (4.3)$$

where

N is the total number of frequency bins k ,

$\mathbf{e} = [1 \quad \mathbf{0}_{2M-2}^T \quad 1]^T$ is the reference microphone selection vector,

\mathbf{w} is the filter weight vector of the beam-forming algorithm,

$\tilde{\mathbf{P}} = \begin{bmatrix} \mathbf{P} & \mathbf{0}_{M \times M} \\ \mathbf{0}_{M \times M} & \mathbf{P} \end{bmatrix}$ is the block diagonal joint CPSD matrix of CPSD \mathbf{P} , of the interferers and the background noise, and

$\tilde{\mathbf{P}}_x$ is the block-diagonal joint CPSD matrix of the CPSD \mathbf{P}_x , of the desired source, defined similar as $\tilde{\mathbf{P}}$.

For the methods proposed, the output noise power can be measured as $\text{Tr}(\mathbf{W}\tilde{\mathbf{P}})$. Hence the SNR^{out} in Eq. (4.3) can be re-written as

$$\text{SNR}^{\text{out}}(l) = \frac{\sum_{k=1}^N \text{Tr}(\mathbf{W}(l, k)\tilde{\mathbf{P}}_x(l, k))}{\sum_{k=1}^N \text{Tr}(\mathbf{W}(l, k)\tilde{\mathbf{P}}(l, k))}. \quad (4.4)$$

The gsSNR can found by averaging the SNR across time-frames as

$$\begin{aligned} \text{gsSNR}^{\text{in}} &= \frac{1}{T} \sum_{l=1}^T 10 \log_{10}(\text{SNR}^{\text{in}}(l)) \quad \text{dB}, \\ \text{gsSNR}^{\text{out}} &= \frac{1}{T} \sum_{l=1}^T 10 \log_{10}(\text{SNR}^{\text{out}}(l)) \quad \text{dB}, \end{aligned} \quad (4.5)$$

where T is the total number of time frames ' l '.

The gain in the gsSNR, which measures the improvement in the quality of desired signal after processing, can be defined by

$$\text{gsSNR}^{\text{gain}} = \text{gsSNR}^{\text{out}} - \text{gsSNR}^{\text{in}} \quad \text{dB}. \quad (4.6)$$

4.2.1.2 Frequency weighted segmental signal-to-noise ratio (fwSegSNR)

As an extension to the global segmental signal-to-noise ratio (gsSNR), by weighting the frequency bands based on their perceptual importance, frequency weighted segmental signal-to-noise ratio (fwSegSNR) can be computed. Here, the fwSegSNR for the left ear, at the input and the output of the beam-former is evaluated using the principle followed in [40] as

$$\text{fwsegSNR}_L^{\text{in}} = \frac{10}{T} \sum_{l=1}^T \frac{\sum_{j=1}^K g(l, j) \log_{10}(\sum_{k \in CB_j} \text{SNR}_L^{\text{in}}(l, k))}{\sum_{j=1}^K g(l, j)}, \quad (4.7)$$

$$\text{fwsegSNR}_L^{\text{out}} = \frac{10}{T} \sum_{l=1}^T \frac{\sum_{j=1}^K g(l, j) \log_{10}(\sum_{k \in CB_j} \text{SNR}_L^{\text{out}}(l, k))}{\sum_{j=1}^K g(l, j)}, \quad (4.8)$$

where

K is the total number of critical bands,

CB_j is the set of frequency bins ‘ k ’, in the j^{th} critical band,

SNR_L^{in} is the monoaural input SNR at the left reference microphone,

$\text{SNR}_L^{\text{out}}$ is the monoaural output SNR at the left HA, and

$g(l, j)$ is the weight placed on the j^{th} frequency band, and the l^{th} time frame.

The critical band SNRs are obtained by dividing the spectrum into 13 critical frequency bands and summing up the SNRs in the frequency bins of each critical band as done in [41]. The weights are chosen based on the spectrum of the clean, desired signal [42].

Similar to Eq. (4.6), the gain in fwSegSNR can given as

$$\text{fwsegSNR}_L^{\text{gain}} = \text{fwsegSNR}_L^{\text{out}} - \text{fwsegSNR}_L^{\text{in}} \quad \text{dB}. \quad (4.9)$$

The fwSegSNR for the right ear can also be found similarly.

4.2.2 Intelligibility Measures

Speech intelligibility refers to the amount of words correctly identified by a listener. Reliable objective intelligibility metrics, with a monotonic relation to the intelligibility of the noise reduction algorithms, help to analyse the performance before executing comprehensive subjective listening tests. In this section, STOI and SIIB are described, which are used to assess the performance of the methods proposed.

4.2.2.1 Short-term objective intelligibility (STOI)

The short-term objective intelligibility (STOI) metric, is a simple, intrusive objective measure, that is suitable to assess the intelligibility of time-frequency weighted speech processing algorithms. STOI generates a scalar value, that is based on the correlation coefficient between the short-time segments of the clean and processed speech [43].

To compute the STOI metric, the temporal envelopes of the clean and processed signal are segmented into time frames of length 386 ms. Then, the correlation coefficient between the clean and the degraded signal is computed per time-segment and frequency bin, and averaged. The **Matlab** implementation made publicly available by the authors of [43], is used to evaluate the metric at each ear.

4.2.2.2 Speech intelligibility in bits (SIIB)

The speech intelligibility in bits (SIIB) metric, is an information theory based intrusive intelligibility metric, that measures the mutual information between the clean and processed signal in bits per second. It is based on the principle that, intelligibility is related to the amount of information that is common between the clean and the processed signal.

In [44], the SIIB metric was shown to generalise better than the STOI metric, especially for speech degraded with modulated point noise sources and reverberation.

Moreover, due to the removal of statistical dependence between the spectro-temporal regions during computation, the intelligibility predicted by the SIIB is more accurate than STOI [45]. Here, the **Matlab** implementation made publicly available by the authors of [44], is used to evaluate the metric for each ear.

4.2.3 Localisation Measures

4.2.3.1 ILD, ITD, ITF Measures

As mentioned previously, binaural cues, specifically the ILD and the ITD enable the auditory brain to localise sound along the horizontal plane. To analyse the preservation of these acoustic cues in the methods proposed, the total errors in the ILD, the ITD and the ITF are evaluated.

As given in Eq. (2.9), Eq. (2.10) and Eq. (2.11), let $\gamma_{u_i}(l, k)$, $\tau_{u_i}(l, k)$ and $\epsilon_{u_i}(l, k)$ be the ITF, the ITD and the ILD errors of the interfering sources at the k^{th} bin and the l^{th} time frame, respectively. In the following evaluations, the assumption that the ILD cues are dominant for $f \geq 1.5$ kHz while the ITD cues are dominant for $f < 1.5$ kHz is held, and errors are measured only in the corresponding frequency ranges [10]. As done in [46], the measures can be evaluated as

$$\text{Total Error}^{\text{ILD}} = \sum_{i=1}^r \left(\frac{1}{N - k_{\text{ILD}} + 1} \sum_{k=k_{\text{ILD}}}^N \left(\frac{1}{T} \sum_{l=1}^T \epsilon_{\mathbf{u}_i}(l, k) \right) \right), \quad (4.10)$$

$$\text{Total Error}^{\text{ITD}} = \sum_{i=1}^r \left(\frac{1}{k_{\text{ILD}} - 1} \sum_{k=1}^{k_{\text{ILD}}-1} \left(\frac{1}{T} \sum_{l=1}^T \tau_{\mathbf{u}_i}(l, k) \right) \right), \quad (4.11)$$

$$\text{Total Error}^{\text{ITF}} = \sum_{i=1}^r \left(\frac{1}{N} \sum_{k=1}^N \left(\frac{1}{T} \sum_{l=1}^T \gamma_{\mathbf{u}_i}(l, k) \right) \right), \quad (4.12)$$

where

r is the total number of interferers considered,

N is the total number of frequency bins k ,

T is the total number of time frames l , and

k_{ILD} is the first frequency bin corresponding to $f \geq 1.5$ kHz.

The error for the target source \mathbf{x} can be computed similarly.

4.3 Simulation Results

In this section, the results of the two methods proposed are compared with the reference methods. The BMVDR algorithm preserves the binaural cues of the target, while it distorts the cues of the interferers. The JBLCMV algorithm preserves the binaural cues of the target and up to r_{max} interfering sources. Since $M = 4$, with the JBLCMV, the spatial cues of up to $r_{\text{max}} = 2M - 3 = 5$ interferers can be preserved. To analyse the performance of the proposed methods, up to $r = 7$ interfering sources are used.

For the R-ILD method, the choice of the parameter c_i is made by comparing the localisation performance against the noise reduction performance for $c_i \in [0, 1]$, in the corresponding environments. As the performance of the two proposed methods are similar in both the anechoic and the reverberant environment, they are discussed together in the next section.

4.3.1 Performance Analysis

For both the anechoic and the office environment, the R-ILD method uses $c_i = c = 0.2$ for all the interferers, as this provides a good trade-off between the ILD cue preservation and the noise reduction, as seen in Figures 4.6a and 4.6b.

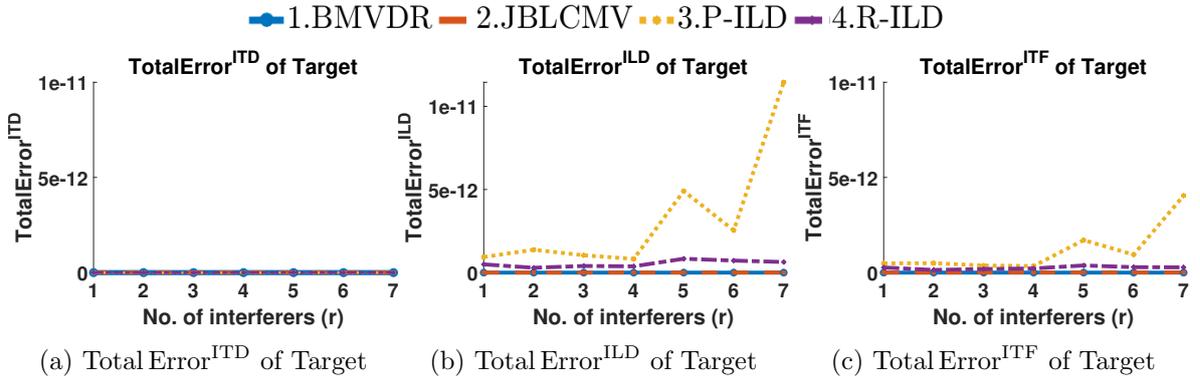


Figure 4.2: Anechoic Environment: Comparing the target source localisation performance of the competing methods in terms of ILD, ITD and ITF.

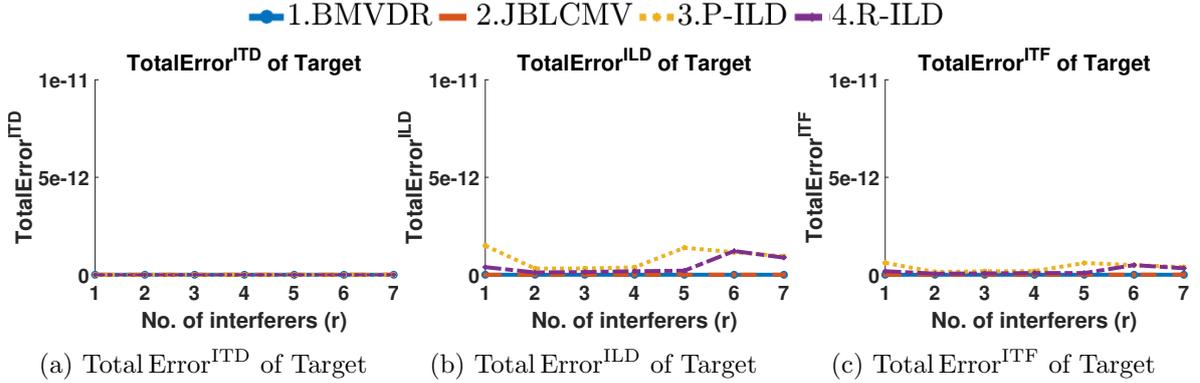


Figure 4.3: Office Environment: Comparing the target source localisation performance of competing methods in terms of ILD, ITD and ITF.

Figures 4.2, 4.3, 4.4 and 4.5 compare the TotalError^{ILD}, TotalError^{ITD} and TotalError^{ITF} of the methods proposed with the reference methods for the binaural cue preservation of the target and the interferers, for both the anechoic and the office environment.

In Figures 4.2 and 4.3, it can be seen that both the ILD and the ITD, and thereby

the ITF, of the target signal are perfectly preserved for all the four methods. This is because, the target is maintained undistorted by all the four methods.

Figures 4.4 and 4.5 compare the performance of the interfering sources. In Section 4.2.3, the ILD and ITD cue errors were defined in the frequencies where they are considered dominant. To assess the cue preservation performance by the proposed methods in detail, the binaural cue errors are evaluated in the complementary frequencies as well. That is, the ILD cue errors are also evaluated for $f < 1.5$ kHz and the ITD cue errors are also evaluated for $f \geq 1.5$ kHz.

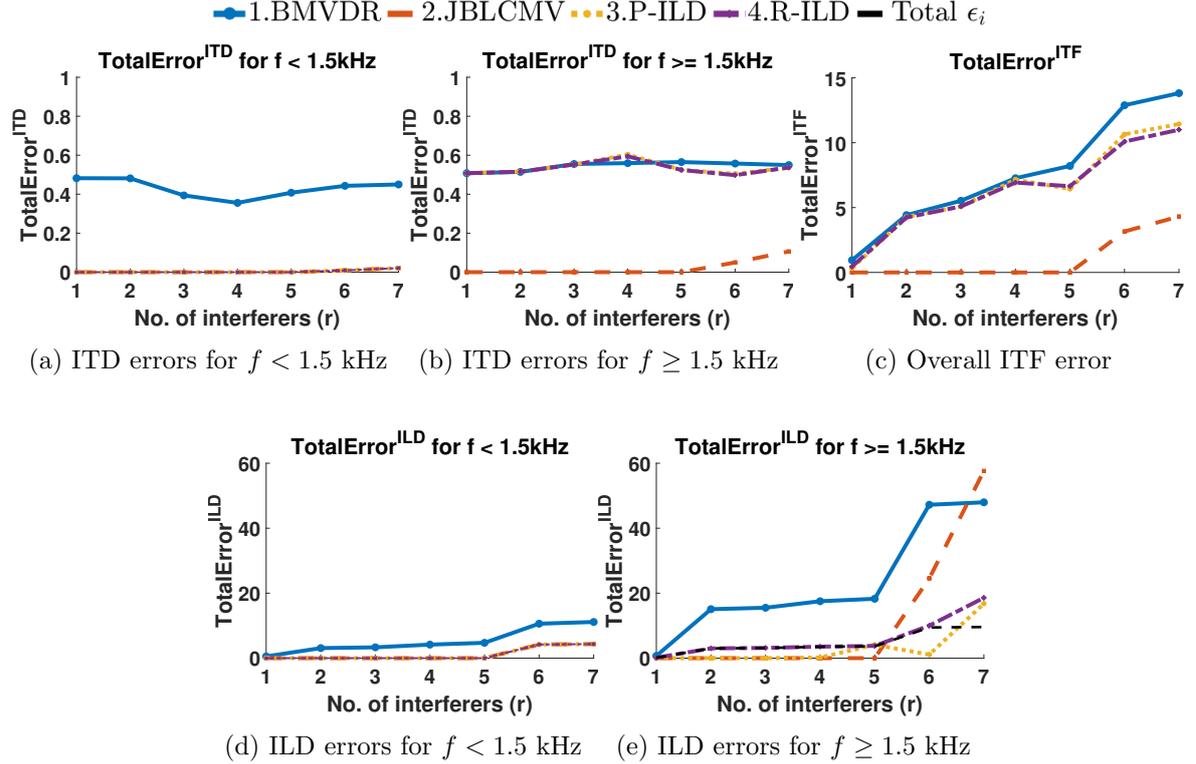


Figure 4.4: Anechoic Environment: Comparing the localisation performance of the interfering sources amongst the competing methods in terms of ILD, ITD and ITF.

For frequencies $f < 1.5$ kHz :

As mentioned in Section 4.1, the JBLCMV algorithm is used by the proposed methods, for $f < 1.5$ kHz. Hence, the $\text{TotalError}^{\text{ITD}}$ and the $\text{TotalError}^{\text{ILD}}$ emulate that of the JBLCMV method as seen in Figures 4.4a, 4.4d and Figures 4.5a, 4.5d.

For frequencies $f \geq 1.5$ kHz :

As seen in Figures 4.4b and 4.5b, the $\text{TotalError}^{\text{ITD}}$ for the P-ILD method and the R-ILD method, are comparable to the errors with the BMVDR algorithm. This is as expected, as both the P-ILD method and the R-ILD method only control the ILD cues, while the ITD cues are unconstrained.

In Figures 4.4e and 4.5e, it can be seen that the $\text{TotalError}^{\text{ILD}}$ by the P-ILD method, is as low as observed with the JBLCMV algorithm for $r \leq 4$, and begins to deviate gradually from $r > 4$. With the R-ILD method, although the $\text{TotalError}^{\text{ILD}}$

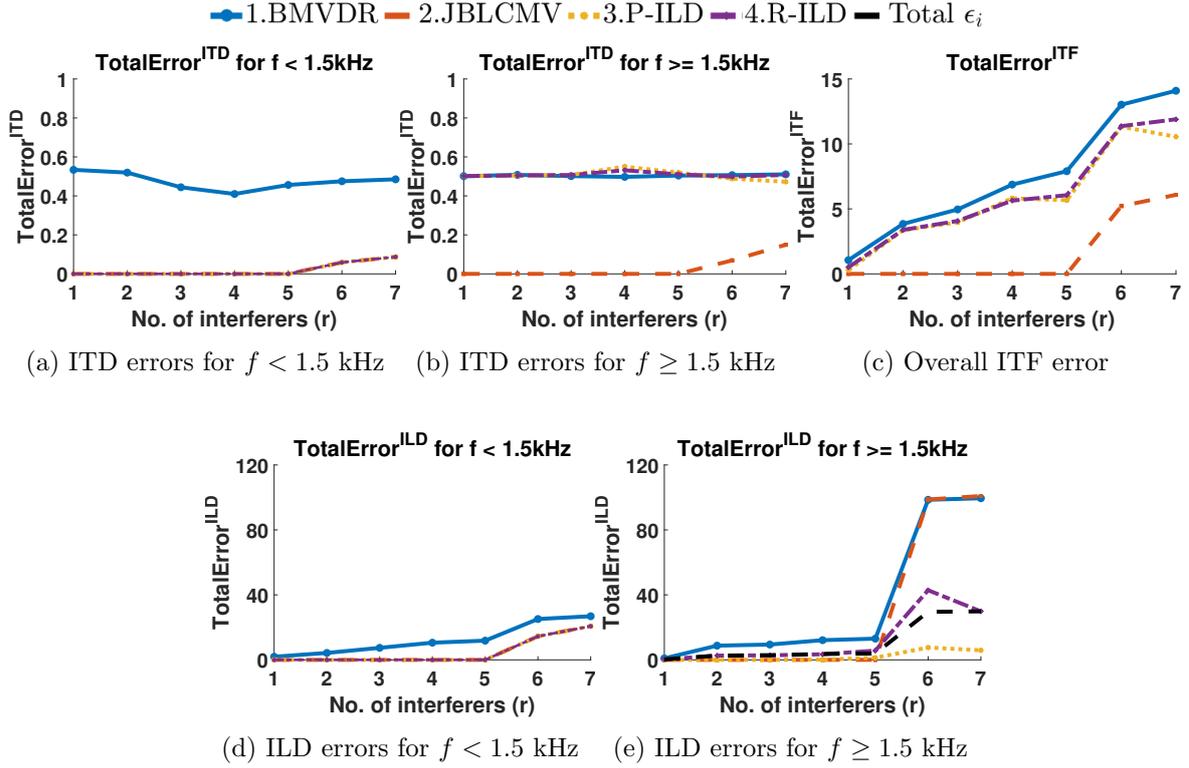


Figure 4.5: Office Environment: Comparing the localisation performance of the interfering sources amongst the competing methods in terms of ILD, ITD and ITF.

error is not perfectly preserved, it can be seen that the error is within Total \mathcal{E}_i , the upper bound specified by Eq. (3.13) for $r \leq 4$, and begins to violate the bound from $r > 4$.

The P-ILD and the R-ILD methods are approximate solutions of the original QCQP problem, as explained in Chapter 3. By increasing the number of interfering sources ‘ r ’ whose ILD cues are to be preserved, the ILD cue constraint violations begin to worsen, as shown in Table 4.2 for the anechoic environment. With the P-ILD method, it can be seen that for $r \leq 3$, the ILD error in every frequency bin is nearly preserved. While, for $r > 3$, the ILD cue error shoots up in certain frequencies. This can be correspondingly observed with the R-ILD method. Thus the Total Error^{ILD} for both the P-ILD method and the R-ILD method increases with ‘ r ’, violating the original ILD cue constraints.

However an interesting behaviour can be noticed for $r > 5$ in Figures 4.4e and 4.5e, where the P-ILD and the R-ILD methods result in lower Total Error^{ILD} than with the JBLCMV. This is because, with the JBLCMV algorithm, the spatial cues for $r > 5$ are completely distorted for every frequency bin, leading to a higher error. Whereas the P-ILD and the R-ILD methods, due to their relaxed problem formulation that uses additional inequality constraints on the ILDs of the interferers, bounds the ILD errors in certain frequency bins as seen in Table 4.2.

Table 4.2: Anechoic Environment: Total Error in ILD over frequency bins.

Number of Interferers	P-ILD	R-ILD
$r = 1$	<p>ILD Error for P-ILD, $r = 1$</p>	<p>ILD Error for R-ILD, $r = 1$</p>
$r = 2$	<p>ILD Error for P-ILD, $r = 2$</p>	<p>ILD Error for R-ILD, $r = 2$</p>
$r = 3$	<p>ILD Error for P-ILD, $r = 3$</p>	<p>ILD Error for R-ILD, $r = 3$</p>
$r = 4$	<p>ILD Error for P-ILD, $r = 4$</p>	<p>ILD Error for R-ILD, $r = 4$</p>
$r = 5$	<p>ILD error for P-ILD, $r = 5$</p>	<p>ILD Error for R-ILD, $r = 5$</p>

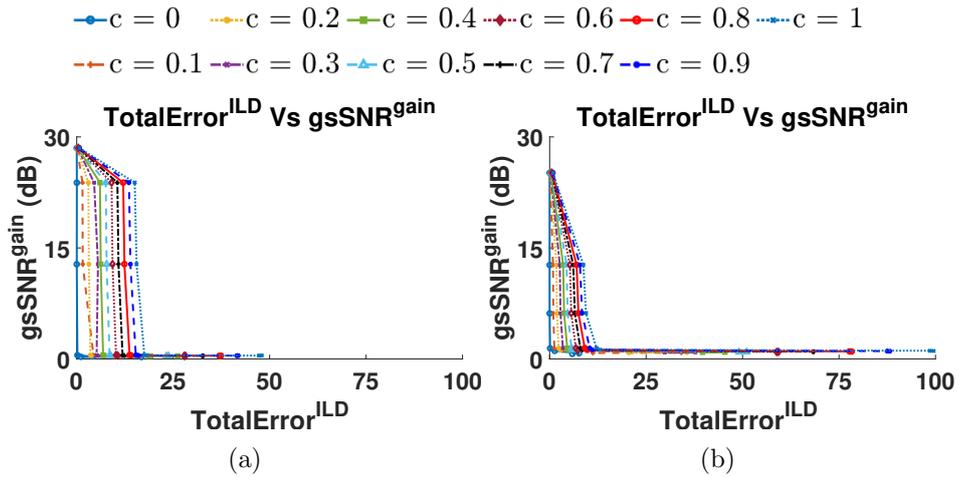
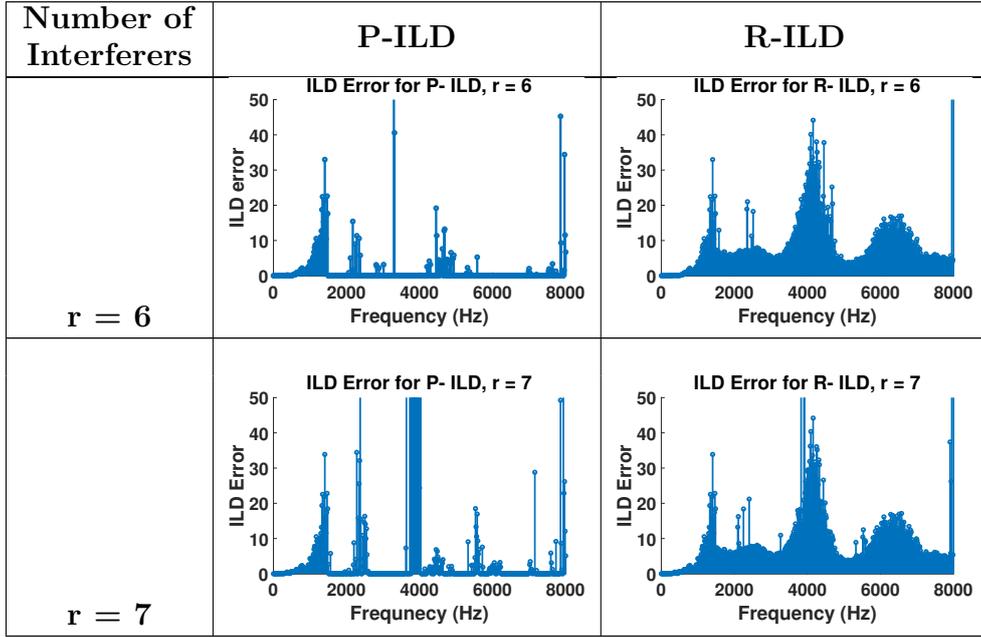


Figure 4.6: Comparing the noise reduction performance against the total ILD error over varying c (a) Anechoic, (b) Office.

In order to choose a suitable ‘ c ’ for the R-ILD method, the effect of ‘ c ’ on noise reduction performance against Total Error^{ILD} was observed as shown in Figures 4.6a and 4.6b. The figures show the performance curves for a different number of simultaneously present sources ‘ r ’ along the curve. It can be seen that the gsSNR^{gain} over different values of ‘ c ’ is nearly the same for a given ‘ r ’. This re-establishes the previous inference, that the level of ILD cue errors does not affect the noise reduction performance of the algorithm, but the number of interfering sources do (due to the increase in constraints). Hence, choosing a lower ‘ c ’ would result in a lower ILD error, up to $r < 4$, and hence, ‘ c ’ is taken to be 0.2 for both, the anechoic and the office environment.

In Figure 4.4c, the TotalError^{ITF} is not preserved with both the P-ILD method

and the R-ILD method. This is because, both methods aim to tackle only the ILD cue errors, while the ITD cues are unconstrained. However, the TotalError^{ITF} is lower than the error with the BMVDR algorithm, due to the lower ILD errors.

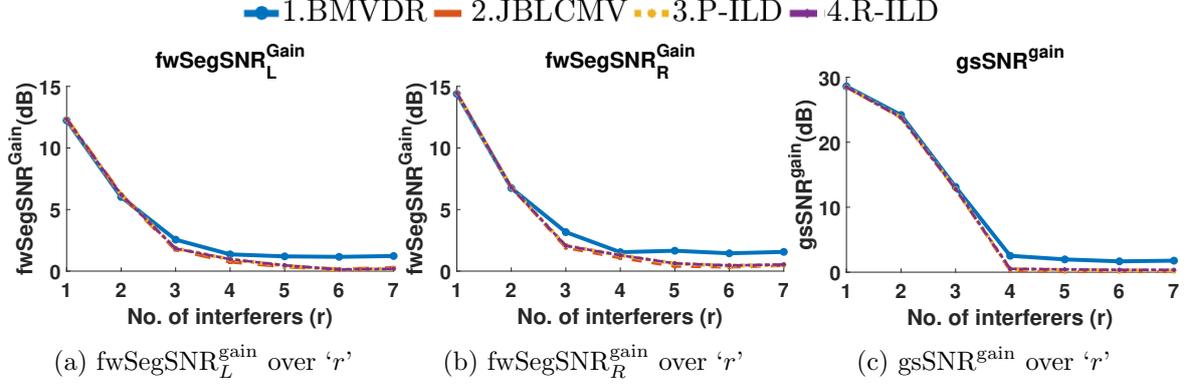


Figure 4.7: Anechoic Environment: Comparing the noise reduction performance of the competing methods in terms of gsSNR and fwSegSNR.

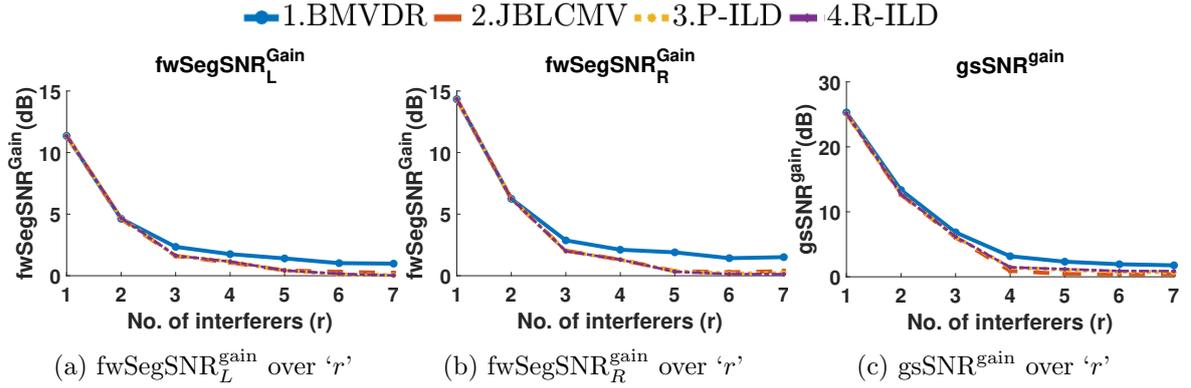


Figure 4.8: Office Environment: Comparing the noise reduction performance of the competing methods in terms of gsSNR and fwSegSNR.

Figures 4.7 and 4.8 compare the noise reduction performance between the proposed methods and the reference methods in terms of the gain in gsSNR and fwSegSNR. It can be seen that the noise reduction with the P-ILD method and the R-ILD method are quite comparable to the noise reduction with the JBLCMV, and lower than that with the BMVDR algorithm. With the P-ILD method, for $r \leq 3$, where the ILD cues are nearly preserved, the noise reduction performance is similar to that with the JBLCMV. Similarly with the R-ILD method, where the ILD errors are bounded, the noise reduction is similar to the JBLCMV. For $r \geq 4$, an improvement of less than 0.5 dB is observed for $\text{gsSNR}^{\text{gain}}$ and $\text{fwSegSNR}^{\text{gain}}$ over the JBLCMV. Here, the modified gsSNR in Eq. (4.4) is used with the P-ILD method and the R-ILD method, which measures the $\text{gsSNR}^{\text{gain}}$ by the approximate methods proposed. As mentioned in Chapter 4, the approximation provides a lower bound, i.e., the noise reduction performance will

be equal to or better than the noise reduction achieved by the original problem. Hence the gain in SNR may be lower than 0.5 dB, when the ILD cues are perfectly preserved. This implies that, preserving the ILD cues or bounding them singularly, instead of preserving both the binaural cues, may provide a very mild gain in SNR, less than 0.5 dB.

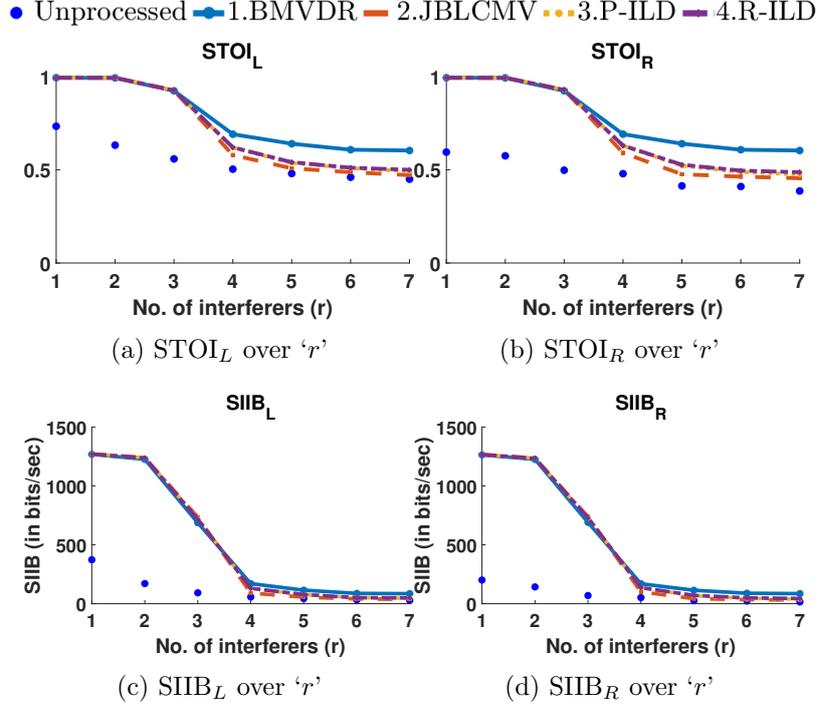


Figure 4.9: Anechoic Environment: Comparing the intelligibility performance of the competing methods in terms of STOI and SIIB.

Figures 4.9 and 4.10 compare the performance in terms of the STOI and the SIIB intelligibility metrics. On observing the measures for $r \leq 3$, where the the P-ILD method nearly preserves the ILD cues, it can be seen that the performance is similar to that with the JBLCMV. Although the R-ILD method does not perfectly preserve the ILD cues even for $r \leq 3$, its intelligibility performance is similar to that observed with the JBLCMV. With increasing $r > 3$, the intelligibility improves mildly with both the P-ILD method and the R-ILD method in comparison to the JBLCMV. With the JBLCMV, the intelligibility is strongly affected by the number of interferers present, and thereby the number of constraints used. While the same applies to the P-ILD method and the R-ILD method, the approximate solution (where the feasibility region is larger) provides a sub-optimal solution, higher in noise power reduction than the true optimal solution, and a small improvement in intelligibility is observed for $r > 3$. Hence by constraining only the ILD cues, the intelligibility improves mildly over a method that preserves both the ILD and the ITD cues perfectly, such as the JBLCMV, when higher number of interferers are present.

Finally, Figure 4.11 compares the localisation performance in terms of

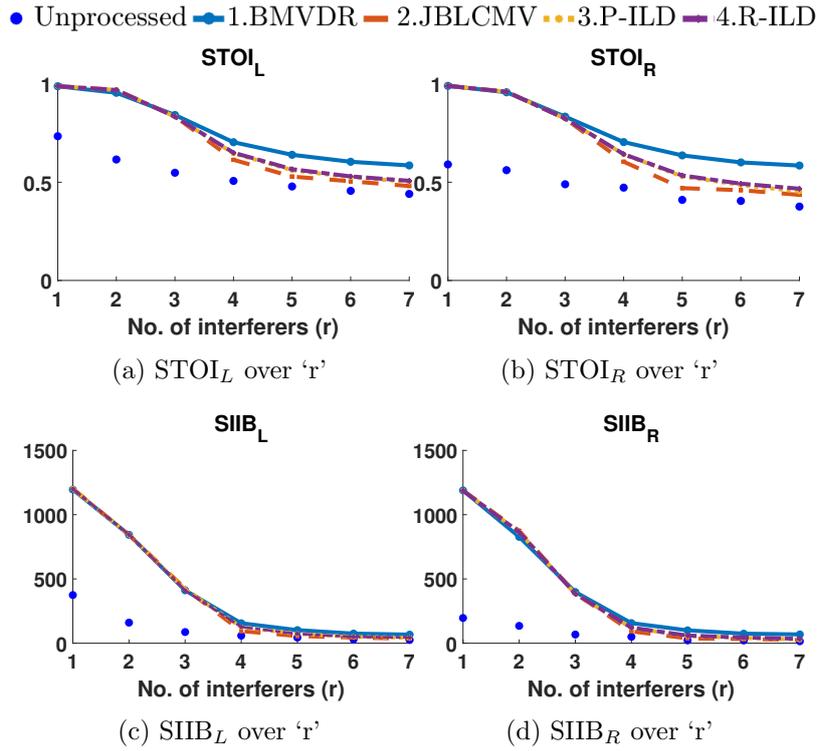


Figure 4.10: Office Environment: Comparing the intelligibility performance of the competing methods in terms of STOI and SIIB.

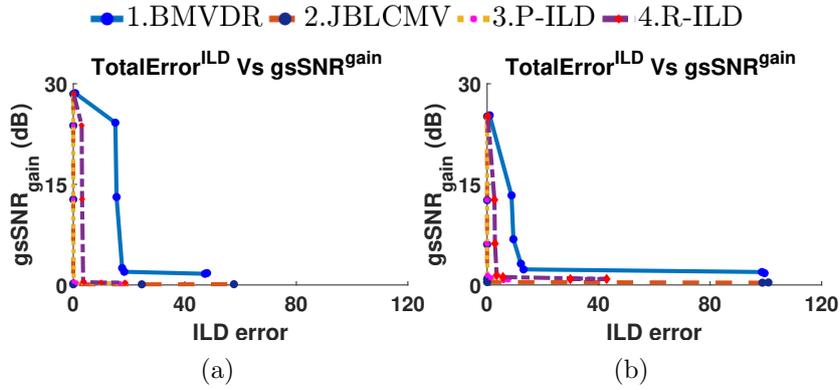


Figure 4.11: Comparing the noise reduction performance against the ILD localisation for the competing methods (a) Anechoic, (b) Office.

Total Error^{ILD}, against the noise reduction performance in terms of gsSNR^{gain}, amongst the competing methods. The figure plots the performance curves for a different number of simultaneously present sources ' r ' along the curve. With the P-ILD method and the R-ILD method, specifically for $r \leq 3$, the noise reduction performance is close to that with the JBLCMV beam-former. For $r > 3$, an improvement in gsSNR^{gain} of about 0.5 dB is observed. Since the gain in noise reduction is an over-estimation, the true gain in noise reduction may be even lower. The Total Error^{ILD} with the P-ILD and the

R-ILD methods are however significantly lower than with the BMVDR. The localisation performance of the P-ILD method is even very similar to that of the JBLCMV for interfering sources $r \leq 4$.

Therefore when the number of interfering source ‘ r ’ is low, the methods proposed preserve the ILD cues without any constraint violations, and there is no gain in noise reduction in comparison to when both the binaural cues are preserved, as done with the JBLCMV method. The improvement in noise reduction is only observed with increasing number of interferers, which also cause errors in localisation performance. This gain in noise reduction is still an over-estimation and hence trying to preserve only the ILD cues, independent of the ITD may only contribute to a small extent, to increasing the DoF available for noise reduction.

4.4 Informal Listening Test

An informal listening test was conducted to analyse the localisation performance, to supplement the simulations performed on the methods proposed. Section 4.4.1 describes the synthesis of the signals and the setup used for the test. The results of the listening test are presented in Section 4.4.2.

4.4.1 Test Setup

To assess the localisation performance of the methods proposed in this thesis against the reference methods, an informal listening test was conducted.

Table 4.3: Listening Test Scenario.

Scene	Target	Interferer 1	Interferer 2	Interferer 3	Microphone Noise	Environment
Signals	Female Speech	Male Speech	Music	HF Signal	WGN	
A	0°	90°	-60°	30°	Yes	Anechoic
B	0°	-75°	-15°	45°	Yes	Office

For the listening tests, the noisy environment contained one desired speech signal, with three interfering point sources and microphone self noise. The tests were performed for anechoic and reverberant office environments, using the BTE HRIRs from [37]. The type of signals used, and the position of the sources are reported in Table 4.3. The target signal is a female speech signal, and Interferer 1 is a male speech signal, taken from the TIMIT database [38]. Interferer 2 is a music signal, while Interferer 3 is a high pass filtered cellphone vibration signal with a cut-off frequency 2 kHz. The signals are uniformly 4 seconds in duration and they are sampled at $f_s = 16$ kHz. The microphone self noise is simulated by additive WGN, such that the target is at a SNR of 50 dB with respect to the WGN. The overall SNR of each scenario is made to be 5 dB. The synthesis of the signals before applying the beam-forming algorithms is done similar to Section 4.1.2.

The subjects were presented with four filtered signals from each of the five algorithms, i.e., BMVDR, JBLCMV, P-ILD, R-ILD($c = 0.1$) and R-ILD($c = 0.3$). Additionally, the subjects were also presented with the clean, unprocessed sources independently, that are taken as the reference position of the sources. This helps to eliminate any bias introduced due to the use of ATF from [37], that are determined using a head and torso simulator (HATS), and are not tailored to the ATF of the individual subject. The test was taken by 20 subjects, with self reported normal hearing. The subjects were in the age group of 23-27 years.

The signals and the scenes were presented in a random order to each subject, with two repetitions. The repetitions of the unprocessed signals are averaged and are taken to be the reference positions. The localisation errors of the subjects are found relative to this reference position and are averaged over their repetitions.

Since the tests could not be conducted in a formal, quiet environment, the tests were taken online by the subjects, and were instructed to emulate a formal setup as much as possible.

4.4.2 Results

Among the results from the 20 subjects who participated in the testing, 2 subjects showed a higher variance in the localisation performance than the rest. Therefore only the results from 18 subjects were considered for the performance analysis.

To assess the localisation performance between the different methods and the sources, a two-way analysis of variance (ANOVA) test is performed, by taking the processing algorithm and the source as the two independent variables. A two-way ANOVA test compares the mean of the localisation errors and indicates if

1. at least two groups have significantly different mean localisation errors with respect to the algorithm used,
2. at least two groups have significantly different mean localisation errors with respect to the source used,
3. there is any interaction between the algorithm used and the source used.

Table 4.4: Two-way ANOVA test for Scene A.

Source of variation	Sum of Squares	DoF	Mean of Squares	F _{value}	p value	F _{critical}
Algorithm	1102.1	3	367.35	0.88	0.45	2.6489
Source	61648	2	30823.99	74.23	0	3.0402
Interaction	1831.5	6	305.25	0.74	0.6219	2.1432
Error	84713.8	204	415.26			
Total	149295.4	215				

Table 4.5: Two-way ANOVA for Scene B.

Source of variation	Sum of Squares	DoF	Mean of Squares	F _{value}	p value	F _{critical}
Algorithm	762.7	3	254.2	0.89	0.4449	2.6489
Source	34575.1	2	17287.6	60.82	0	3.0402
Interaction	403.1	6	67.2	0.24	0.9642	2.1432
Error	57984.3	204	284.2			
Total	93725.2	215				

Tables 4.4 and 4.5 show the results of a two-way ANOVA test conducted for the JBLCMV, P-ILD, R-ILD($c = 0.1$) and R-ILD($c = 0.3$), for the four sources at a 0.95 confidence level. Across the algorithms, it can be seen that the $F_{\text{value}} < F_{\text{critical}}$, thereby indicating that the mean values of the localisation errors are equal across the algorithms considered. Across the sources used, the $F_{\text{value}} > F_{\text{critical}}$, indicating that at least two sources have a different mean localisation error. Finally, it can also be observed that, there is no significant interaction between the source used and the algorithm applied. This implies that the algorithms behave similarly in terms of localisation, independent of the source used.

Table 4.6: T-test p-values for Scene A.

Algorithm	BMVDR	JBLCMV	P-ILD	R-ILD($c = 0.1$)	R-ILD($c = 0.3$)
P-ILD	2.1950e-14	0.1876	1	0.4654	0.0622
R-ILD($c = 0.1$)	5.5470e-14	0.0872	0.4654	1	0.1808
R-ILD($c = 0.3$)	1.2763e-12	0.0025	0.0622	0.1808	1

Table 4.7: T-test p-values for Scene B.

Algorithm	BMVDR	JBLCMV	P-ILD	R-ILD($c = 0.1$)	R-ILD($c = 0.3$)
P-ILD	1.4368e-13	0.7691	1	0.3088	0.0924
R-ILD($c = 0.1$)	1.5553e-14	0.1682	0.3088	1	0.3288
R-ILD($c = 0.3$)	6.9744e-14	0.0591	0.0924	0.3288	1

To compare the algorithms proposed with the reference methods individually, paired T-tests at 0.95 confidence level were also performed for each scene. The ‘p-values’ of the T-tests are shown in Tables 4.6 and 4.7. In both Scene A and Scene B, the methods proposed localise significantly better than with the BMVDR method. It can be seen that in Scene A, the P-ILD and the R-ILD($c = 0.1$) methods are not significantly different in localisation performance, to the JBLCMV method. In Scene B, the P-ILD localisation performance is very similar to the JBLCMV and the R-ILD($c = 0.1$) and the R-ILD($c = 0.3$) are not significantly different from the JBLCMV. On comparing amongst the methods proposed, it can be seen that the P-ILD method performs better or similar to the R-ILD($c = 0.1$) method in most cases, and as expected, the R-

ILD($c = 0.3$) method performs poorer by comparison.

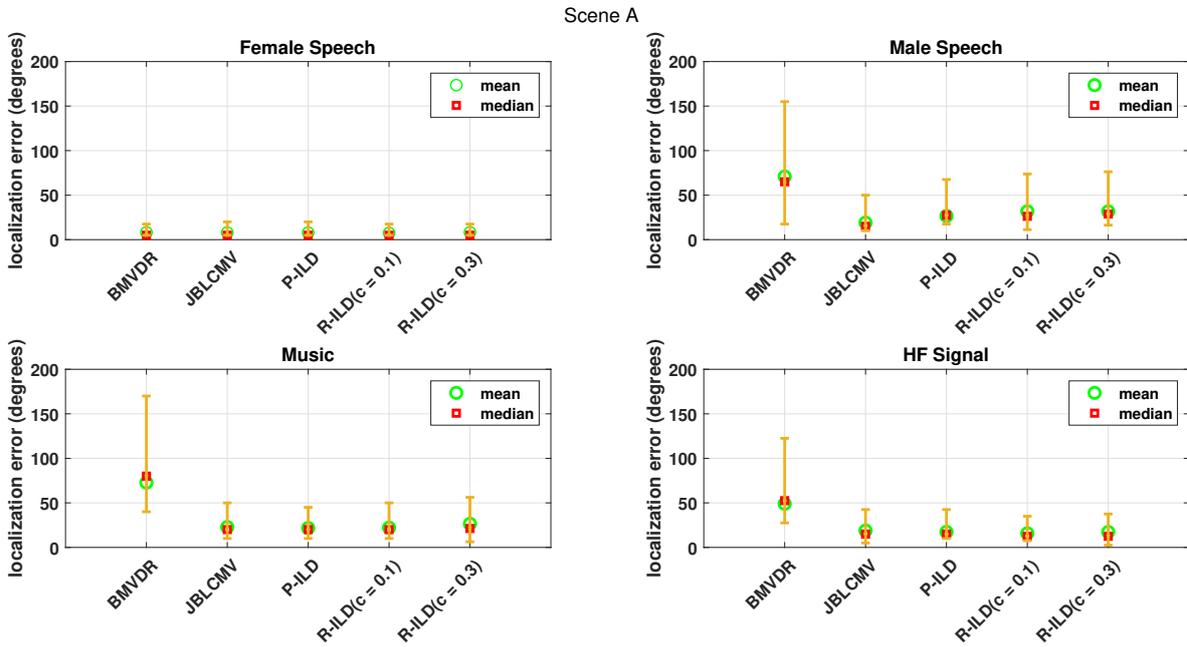


Figure 4.12: Anechoic Environment: Localisation error for each source.

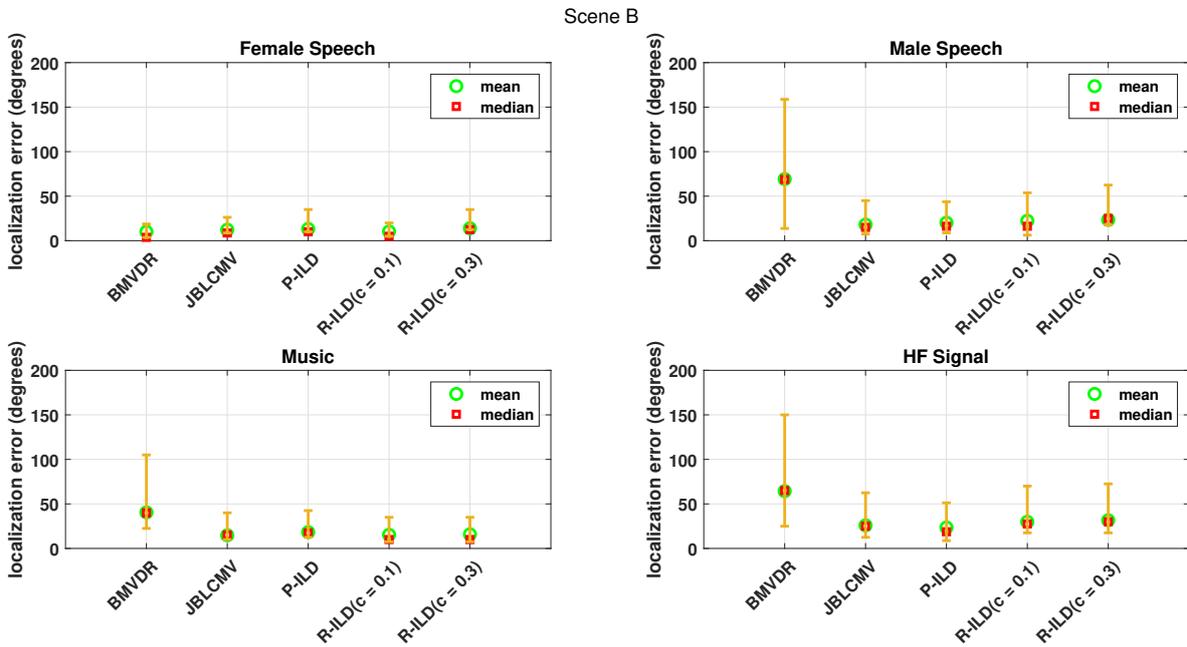


Figure 4.13: Office Environment: Localisation error for each source.

Since the localisation performance was found to vary among the sources, the statis-

tics of the localisation error for each source is calculated, as shown in Figures 4.12 and 4.13. The figures show the 0.25, 0.75 quartiles, the mean and median of the localisation errors for each algorithm and source.

As expected, the target female speech signal shows the least localisation error for all the five algorithms, as the target is constrained to be distortionless. Moreover, the interfering signals show the maximum localisation error with the BMVDR algorithm. For the interfering sources, it can be seen that the localisation errors by the methods proposed is not significantly different from the errors observed with the JBLCMV method. Comparing the performance of the three interfering sources, in Scene A it can be seen that the 0.75 quartile of the localisation errors for the male speech signal is higher than that observed for the other two sources. Since both music and the male speech signal cover the entire frequency spectrum, the anomaly in the male speech source can be attributed to the non-uniform testing conditions among the subjects.

Overall, the listening tests showed that there is no significant difference in the localisation performance between the JBLCMV method and the P-ILD and R-ILD methods, for the three interfering sources. This implies that the localisation performance of the interferers, when only the ILD cues are preserved for the higher frequencies, matches the performance when both the ILD and the ITD cues are preserved. Moreover, although the ILD cues are not perfectly preserved with the R-ILD method, as seen in Figures 4.4e and 4.5e, the mild ILD cue distortions do not perceptually affect the localisation performance with smaller values of ' c '.

The objective of the thesis was to answer the question: ‘Will beam-forming with only the dominant binaural cue preservation of the noise components, help to improve the noise reduction performance, as opposed to the preservation of both the interaural time difference (ITD) and the interaural level difference (ILD) cues?’. In a quest to answer the question, a first step was taken in this thesis by considering only the ILD cues, the cue dominant in the higher frequencies.

To preserve only the ILD cues, firstly, the noise reduction optimisation problem was re-formulated. In addition to the constraints that keep the target signal undistorted, new constraints to preserve only the ILD cues of the noise components were introduced. In Chapter 3, two methods of constraining the ILD cues were proposed. The first method preserves the ILD cues perfectly, while the second method bounds the ILD cue errors within a suitable limit. Since both the methods had non-convex formulations, a convex relaxation of the problem as a semi-definite program (SDP) was proposed, that provided approximate solutions to the original non-convex formulation.

Secondly, the formulations were experimentally tested, to understand the effect of the ILD cue preservation on the noise reduction performance. Chapter 4 described the results obtained through simulations, and compared the two methods proposed, with the JBLCMV and the BMVDR as the reference methods. The simulations were performed for an anechoic environment and a reverberant environment, and the results for the both were found to be similar. Additionally, informal listening tests were conducted to analyse the localisation performance of the methods proposed.

5.1 Discussion

5.1.1 Method 1: P-ILD

In the first method proposed, the problem was formulated such that, the ILD cues of the noise components were to be preserved perfectly. This introduced one quadratic equality constraint per interferer, making it a non-convex quadratically constrained quadratic program (QCQP) formulation. A convex relaxation of the problem was then formulated, that approximates the original non-convex QCQP well. On performing experiments as done in Chapter 4, it was found that the relaxed convex SDP, resulted in a good approximation when the number of interferers are lower. By increasing the number of interfering sources considered, the violations of the ILD cue preservation constraints increased, and led to a poor approximation. Nevertheless, the optimal solution of the relaxed SDP, provides a non-trivial lower bound to the optimal solution

of the original QCQP. This helped to shed light on the noise reduction ability of this formulation, against the JBLCMV and the BMVDR beam-former.

5.1.2 Method 2: R-ILD

In the second method, a problem formulation that enforced an upper bound to the ILD cue errors of the noise components, using inequality constraints was introduced. The upper bound was selected to be a fraction ‘ c ’, of the ILD error found with the BMVDR beam-former. This formulation was inspired by the work done in [26]. The factor ‘ c ’ was chosen by comparing the noise reduction performance against the ILD cue preservation. This method introduces two non-convex inequality constraints per interfering source. Such a formulation also resulted in a non-convex QCQP. Hence the approximate SDP formulation, as done with the P-ILD was used here. Once again, it was found to be a good approximation when the number of interferers are low, while the constraint violations worsened on increasing the number of interfering sources.

The noise reduction performance of the methods proposed here were compared against the JBLCMV and the BMVDR beam-formers, using simulations. For the simulations, the ILD cues of the interferers were constrained for the higher frequencies using the P-ILD and the R-ILD methods, while the JBLCMV beam-former was used for the lower frequencies. On performing the experiments, in Chapter 4, it was shown that the noise reduction ability of the newly proposed methods, improves by less than 0.5 dB in SNR, over that with the JBLCMV beam-former, where both the ITD and the ILD are controlled. Moreover, on increasing the number of interfering sources, the intelligibility was also seen to be mildly better than with the JBLCMV beam-former. This improvement in the noise reduction and intelligibility was observed for $r > 4$, where violations in the ILD cue preservation occurred. Hence the optimal noise reduction by the original problem may be equal to or lower than that observed with the approximate QCQP. Nevertheless, the approximations were able to provide a non-trivial inference. Moreover, the informal listening tests conducted in Chapter 4, validated the ILD cue preservation performance of the methods proposed and also showed that the localisation performance was good even with bounded errors in the ILD preservation.

In conclusion, by constraining only the ILD spatial cues of the interferers in the higher frequencies, for a small number of interferers, the methods proposed in this thesis show a good localisation performance and a mild improvement in the noise reduction performance. The DoF available for the noise reduction are however, more significantly affected by the number of interfering sources present in the environment.

5.2 Recommendations

A few ideas, that can prove to be useful extensions to the work proposed in this thesis are recommended here.

- **Preserving perceptually audible interferers**

In the methods proposed in this thesis, it was seen that the approximation works well when the number of interfering sources considered are fewer in number. In [25], a method to only constrain perceptually audible interferers for binaural cue preservation is proposed. Perhaps, by constraining only those interferers that are perceptually audible after processing, and thereby reducing the number of interferers constrained, may allow a reliable ILD cue preservation of the interferers that are audible.

- **Preservation of the ILD cues based on the position of the source**

It was seen that by increasing the number of interferers, the ILD cue errors began to shoot up in certain frequency bins. It has been proven in [47], that the sensitivity to the ILD change, reduces as the position of the source moves away from the mid-line (in the horizontal plane) of the listener, in the higher frequencies. It would be interesting to observe, through listening experiments, if allowing larger ILD cue errors in positions away from the mid-line would enable the listener maintain their localisation ability.

- **Preservation of the ITD cues in the lower frequencies**

In this thesis, the focus was on preserving the ILD cues for the higher frequencies, where the ILD cues are dominant. However, for the lower frequencies, the ITD cues dominate over the ILD cues. Hence, to faithfully answer the original research question, similar to preserving the ILD cues alone for the higher frequencies, the ITD cues for the lower frequencies alone must be preserved. Hence, to check if the inference corroborates well when only the ITD cues are preserved for the lower frequencies, a formulation for the preservation of the ITD cues was attempted and can be given by

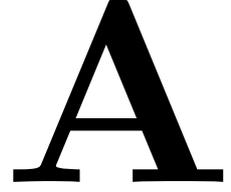
$$\begin{aligned}
& \underset{\mathbf{w}_L, \mathbf{w}_R, \eta > 0}{\text{minimise}} && \begin{bmatrix} \mathbf{w}_L^H & \mathbf{w}_R^H \end{bmatrix} \tilde{\mathbf{P}} \begin{bmatrix} \mathbf{w}_L \\ \mathbf{w}_R \end{bmatrix} \\
& \text{subject to} && \mathbf{w}_L^H \mathbf{a} = a_1, \quad \mathbf{w}_R^H \mathbf{a} = a_M, \\
& && \mathbf{w}_L^H \mathbf{b}_i b_{i,M} - \eta_i \mathbf{w}_R^H \mathbf{b}_i b_{i,1} = 0, \quad i = 1, \dots, r \leq r_{max}, \\
& && \mathbf{w}_L, \mathbf{w}_R \in \mathbb{C}^{M \times 1}, \quad \eta_i \in \mathbb{R}_{++}.
\end{aligned} \tag{5.1}$$

The idea behind Eq. (5.1), is to let the IPD cues be preserved, by letting the output ILD cues vary due to the real valued variable η_i . Using this formulation for the lower frequencies, and using the methods proposed in this thesis for the higher frequencies, the idea of preserving only the dominant cues can be achieved.

- **Estimation of the ATF vectors**

Throughout the thesis, the simulations were conducted using true ATF vectors, to avoid steering vector mismatches. Another approach that could be tested with, would be to estimate the ATF vectors, and observe the performance with respect to localisation due to the steering vector mismatch.

Mathematical Properties



A.1 Linear Algebra

$$\mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} = \text{Tr} \left(\mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \right) = \text{Tr} \left(\mathbf{w} \mathbf{w}^H \tilde{\mathbf{P}} \right) \quad (\text{A.1})$$

A.2 Schur's Complement & Positive Definiteness

Consider a matrix $X \in \mathbf{S}^n$ partitioned as

$$X = \begin{bmatrix} A & B \\ B^H & C \end{bmatrix},$$

where $A \in \mathbf{S}^k$, If $\det A \neq 0$, the matrix

$$S = C - B^H A^{-1} B,$$

is the Schur's complement of A in X [31]. The definiteness of X can found as

- $X \succ 0$ if and only if $A \succ 0$ and $S \succ 0$
- If $A \succ 0$, then $X \succeq 0$ if and only if $S \succeq 0$

For a more generalised case, where A is singular

$$S = C - B^H A^\dagger B$$

$$X \succeq 0 \iff A \succeq 0, \quad (I - AA^\dagger) B = 0, \quad S \succeq 0. \quad (\text{A.2})$$

A.3 Lower bound of SDR

The non-convex QCQP is

$$\begin{aligned} (\text{P}) \quad & \underset{\mathbf{w} \in \mathbb{C}^{2M \times 1}}{\text{minimise}} \quad \mathbf{w}^H \tilde{\mathbf{P}} \mathbf{w} \\ & \text{subject to} \quad \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\ & \quad \mathbf{w}^H \mathbf{M}_i \mathbf{w} = 0, \quad i = 1, \dots, r \leq r_{max}. \end{aligned} \quad (\text{A.3})$$

The dual of Eq.(A.3) can be written as

$$\begin{aligned} (\text{D}) \quad & \underset{\nu, \mu}{\text{maximise}} \quad -\gamma \\ & \text{subject to} \quad \begin{bmatrix} \left(\tilde{\mathbf{P}} + \sum_{i=1}^r \mu_i \mathbf{M}_i \right) & \boldsymbol{\Lambda}_A^H \nu \\ \nu^H \boldsymbol{\Lambda}_A & \nu^H \mathbf{f}_A + \gamma \end{bmatrix} \succeq 0, \end{aligned} \quad (\text{A.4})$$

where ν, μ are the Lagrangian multipliers of the equality constraints of primal problem **P**.

The dual of the dual in Eq.(A.4) is

$$\begin{aligned}
(\mathbf{SDR}) \quad & \underset{\mathbf{w}, \mathbf{W}}{\text{minimise}} \quad \text{Tr}(\mathbf{W}\tilde{\mathbf{P}}) \\
& \text{subject to } \mathbf{w}^H \boldsymbol{\Lambda}_A = \mathbf{f}_A^H \\
& \quad \text{Tr}(\mathbf{W}\mathbf{M}_i) = 0, \quad i = 1, \dots, r \leq r_{max} \\
& \quad \begin{bmatrix} \mathbf{W} & \mathbf{w} \\ \mathbf{w}^H & 1 \end{bmatrix} \succeq 0.
\end{aligned} \tag{A.5}$$

Since the **D** is the dual of **P**, it is a convex problem despite the non-convexity of problem **P**. Moreover, strong duality holds between the problem **D** and its dual **SDR** (Slater's constraint qualifications are satisfied) [31]. Hence the optimal solution of the problem **SDR** is a lower bound to the optimal solution of the problem **P**.

Bibliography

- [1] H. Levitt, “Noise reduction in hearing aids: A review,” *Journal of rehabilitation research and development*, vol. 38, no. 1, pp. 111–122, 2001.
- [2] J. M. Kates, *Digital hearing aids*. Plural publishing, 2008.
- [3] T. Lotter, *Single and Multimicrophone Speech Enhancement for Hearing Aids*. PhD thesis, RWTH Aachen, Germany, Aug. 2004.
- [4] B. D. Van Veen and K. M. Buckley, “Beamforming: a versatile approach to spatial filtering,” *IEEE ASSP Magazine*, vol. 5, no. 2, pp. 4–24, 1988.
- [5] J. G. Desloge, W. M. Rabinowitz, and P. M. Zurek, “Microphone-array hearing aids with binaural output. i. fixed-processing systems,” *IEEE Transactions on Speech and Audio Processing*, vol. 5, no. 6, pp. 529–542, 1997.
- [6] N. W. MacKeith and R. R. A. Coles, “Binaural advantages in hearing of speech,” *The Journal of Laryngology Otology*, vol. 85, no. 3, p. 213–232, 1971.
- [7] A. W. Bronkhorst, “The cocktail party phenomenon: A review of research on speech intelligibility in multiple-talker conditions,” *Acustica*, vol. 86, pp. 117–128, 2000.
- [8] R. Gilkey and T. R. Anderson, *Binaural and spatial hearing in real and virtual environments*. Hillsdale, NJ: Lawrence Erlbaum Associates, Inc., 1997.
- [9] D. Wang and G. J. Brown, *Computational Auditory Scene Analysis: Principles, Algorithms, and Applications*. Wiley-IEEE Press, 2006.
- [10] F. L. Wightman and D. J. Kistler, “The dominant role of low-frequency interaural time differences in sound localization,” *The Journal of the Acoustical Society of America*, vol. 91, pp. 1648–1661, Mar. 1992.
- [11] J. Strutt and L. Rayleigh, “On the perception of sound direction,” *Philos. Mag*, vol. 13, pp. 214–232, 1907.
- [12] E. A. Macpherson and J. C. Middlebrooks, “Listener weighting of cues for lateral angle: the duplex theory of sound localization revisited,” *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2219–2236, 2002.
- [13] J. Thiemann, M. Müller, D. Marquardt, S. Doclo, and S. van de Par, “Speech enhancement for multimicrophone binaural hearing aids aiming to preserve the spatial auditory scene,” *EURASIP Journal on Advances in Signal Processing*, vol. 2016, no. 1, p. 12, 2016.
- [14] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, “Theoretical analysis of linearly constrained multi-channel wiener filtering algorithms for combined noise reduction and binaural cue preservation in binaural hearing aids,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2384–2397, 2015.

- [15] S. Doclo, S. Gannot, M. Moonen, A. Spriet, S. Haykin, and K. R. Liu, “Acoustic beamforming for hearing aid applications,” *Handbook on array processing and sensor networks*, pp. 269–302, 2010.
- [16] B. Cornelis, S. Doclo, T. Van dan Bogaert, M. Moonen, and J. Wouters, “Theoretical analysis of binaural multimicrophone noise reduction techniques,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 18, no. 2, pp. 342–355, 2009.
- [17] T. Van den Bogaert, J. Wouters, S. Doclo, and M. Moonen, “Binaural cue preservation for hearing aids using an interaural transfer function multichannel wiener filter,” in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP’07*, vol. 4, pp. IV–565, IEEE, 2007.
- [18] S. Doclo, R. Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen, “Extension of the multi-channel wiener filter with itd cues for noise reduction in binaural hearing aids,” in *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, 2005.*, pp. 70–73, IEEE, 2005.
- [19] S. Doclo, R. Dong, T. J. Klasen, J. Wouters, S. Haykin, and M. Moonen, “Extension of the multi-channel wiener filter with localisation cues for noise reduction in binaural hearing aids,” in *Proc. IWAENC*, pp. 221–224, 2005.
- [20] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, “Optimal binaural lcmv beamformers for combined noise reduction and binaural cue preservation,” in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 288–292, IEEE, 2014.
- [21] E. Hadad, D. Marquardt, S. Doclo, and S. Gannot, “Theoretical analysis of binaural transfer function mvdr beamformers with interference cue preservation constraints,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 23, no. 12, pp. 2449–2464, 2015.
- [22] E. Hadad, S. Gannot, and S. Doclo, “Binaural linearly constrained minimum variance beamformer for hearing aid applications,” in *IWAENC 2012; International Workshop on Acoustic Signal Enhancement*, pp. 1–4, VDE, 2012.
- [23] A. I. Koutrouvelis, R. C. Hendriks, J. Jensen, and R. Heusdens, “Improved multimicrophone noise reduction preserving binaural cues,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 460–464, IEEE, 2016.
- [24] A. I. Koutrouvelis, J. Jensen, M. Guo, R. C. Hendriks, and R. Heusdens, “Binaural speech enhancement with spatial cue preservation utilising simultaneous masking,” in *2017 25th European Signal Processing Conference (EUSIPCO)*, pp. 598–602, IEEE, 2017.
- [25] C. A. Kokke, R. C. Hendriks, and A. I. Koutrouvelis, “Binaural beamforming based on automatic interferer selection,” in *ICASSP 2019-2019 IEEE International*

- Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 6850–6854, IEEE, 2019.
- [26] A. Koutrouvelis, *Multi-microphone noise reduction for hearing assistive devices*. PhD thesis, Delft University of Technology, 2018.
- [27] D. Marquardt, E. Hadad, S. Gannot, and S. Doclo, “Optimal binaural lcmv beamformers for combined noise reduction and binaural cue preservation,” in *2014 14th International Workshop on Acoustic Signal Enhancement (IWAENC)*, pp. 288–292, IEEE, 2014.
- [28] S. Sahni, “Computationally related problems,” *SIAM Journal on computing*, vol. 3, no. 4, pp. 262–279, 1974.
- [29] A. Beck and Y. C. Eldar, “Strong duality in nonconvex quadratic optimization with two quadratic constraints,” *SIAM Journal on optimization*, vol. 17, no. 3, pp. 844–860, 2006.
- [30] Y. Hsia, G.-X. Lin, R.-L. Sheu, *et al.*, “A revisit to quadratic programming with one inequality quadratic constraint via matrix pencil,” *PACIFIC JOURNAL OF OPTIMIZATION*, vol. 10, no. 3, pp. 461–481, 2014.
- [31] S. Boyd, S. P. Boyd, and L. Vandenberghe, *Convex optimization*. Cambridge university press, 2004.
- [32] Z.-Q. Luo, W.-K. Ma, A. M.-C. So, Y. Ye, and S. Zhang, “Semidefinite relaxation of quadratic optimization problems,” *IEEE Signal Processing Magazine*, vol. 27, no. 3, pp. 20–34, 2010.
- [33] J. Park and S. Boyd, “General heuristics for nonconvex quadratically constrained quadratic programming,” 2017.
- [34] L. Vandenberghe and S. Boyd, “Semidefinite programming,” *SIAM review*, vol. 38, no. 1, pp. 49–95, 1996.
- [35] H. D. Sherali and W. P. Adams, *A reformulation-linearization technique for solving discrete and continuous nonconvex problems*, vol. 31. Springer Science & Business Media, 2013.
- [36] M. Grant and S. Boyd, “CVX: Matlab software for disciplined convex programming, version 2.1.” <http://cvxr.com/cvx>, Mar. 2014.
- [37] H. Kayser, S. D. Ewert, J. Anemüller, T. Rohdenburg, V. Hohmann, and B. Kollmeier, “Database of multichannel in-ear and behind-the-ear head-related and binaural room impulse responses,” *EURASIP Journal on Advances in Signal Processing*, vol. 2009, no. 1, p. 298605, 2009.
- [38] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, D. S. Pallett, N. L. Dahlgren, and V. Zue, “Timit acoustic-phonetic continuous speech corpus,” *Linguistic Data Consortium*, 1993.

- [39] J. Benesty, M. M. Sondhi, and Y. Huang, *Springer handbook of speech processing*. Springer, 2007.
- [40] Y. Hu and P. C. Loizou, “Evaluation of objective quality measures for speech enhancement,” *IEEE Transactions on audio, speech, and language processing*, vol. 16, no. 1, pp. 229–238, 2007.
- [41] P. C. Loizou, *Speech enhancement: theory and practice*. CRC press, 2013.
- [42] J. Ma, Y. Hu, and P. C. Loizou, “Objective measures for predicting speech intelligibility in noisy conditions based on new band-importance functions,” *The Journal of the Acoustical Society of America*, vol. 125, no. 5, pp. 3387–3405, 2009.
- [43] C. H. Taal, R. C. Hendriks, R. Heusdens, and J. Jensen, “An algorithm for intelligibility prediction of time–frequency weighted noisy speech,” *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 7, pp. 2125–2136, 2011.
- [44] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, “An instrumental intelligibility metric based on information theory,” *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 115–119, 2017.
- [45] S. Van Kuyk, W. B. Kleijn, and R. C. Hendriks, “An evaluation of intrusive instrumental intelligibility metrics,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 26, no. 11, pp. 2153–2166, 2018.
- [46] A. I. Koutrouvelis, R. C. Hendriks, R. Heusdens, and J. Jensen, “Relaxed binaural lcmv beamforming,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 25, no. 1, pp. 137–152, 2016.
- [47] W. A. Yost and R. H. Dye Jr, “Discrimination of interaural differences of level as a function of frequency,” *The Journal of the Acoustical Society of America*, vol. 83, no. 5, pp. 1846–1851, 1988.