Beamforming and Blind Signal Separation for Far-field Voice Capture using a Microphone Array

JOHAN DAVIDSSON
LUKE POSTEMA
MASTER´S THESIS
DEPARTMENT OF ELECTRICAL AND INFORMATION TECHNOLOGY
FACULTY OF ENGINEERING | LTH | LUND UNIVERSITY

LUND UNIVERSITY, FACULTY OF ENGINEERING

MASTER THESIS

Beamforming and Blind Signal Separation for Far-field Voice Capture using a Microphone Array

Authors:
Johan DAVIDSSON
dic14jda@student.lu.se

Luke POSTEMA
elt14lpo@student.lu.se

Supervisors:
Markus TÖRMÄNEN (LTH)
markus.tormanen@eit.lth.se

Danny SMITH (Axis)
danny.smith@axis.com

In collaboration with

Axis Communications AB

February 7, 2019

Abstract

With the evolving technology of mobile electronics and other forms of communication methods, an increasing demand of speech intelligibility is introduced. In a conference room for example, multiple microphones are placed directly in front of the speaker in order to pick up clean speech. However, several challenges still exist. For instance the distance between speaker and microphone will not be fixed, unwanted noise from interfering sources around the room might be picked up and reverberation from the actual signal of interest might be introduced. The same problem exists whilst talking into a mobile device when the wind is blowing in the background, which might be easier to relate to. There is a need for adaptive methods which takes these parameters in to consideration in order to improve speech intelligibility.

One way to solve this problem is by the use of an acoustic beam. Ideally, this will leave the source of interest untouched and suppress all unwanted noise. This can be done either by putting a beam in a predetermined direction of arrival or adaptable direction, commonly known as blind signal separation. This is achieved with help of multiple microphones working in tandem, hence the title *Beamforming and Blind Signal Separation for Far-field Voice Capture using a Microphone Array*.

This thesis will investigate the possibility of using beamforming for far-field voice capture using a commercially available microphone array. The concepts are explained and then one direction-of-arrival and two beamforming algorithms are implemented, tested and evaluated. The direction-of-arrival algorithm is based on cross correlation and the beamforming algorithms are known as *delay-and-sum* and *minimum variance distortionless response*.

Acknowledgements

To begin with, we would like to thank Danny Smith, our supervisor at Axis Communications AB, for his help and support throughout the course of this thesis. We also wish to thank Markus Törmänen and Pietro Andreani from Lund University, Faculty of Engineering, for being our supervisor and examiner, respectively. We would also like to thank Axis for this great opportunity and learning experience and our colleagues in the audio team for showing such great interest in our work. Last but not least, our sincere gratitude to Santhosh Nadig for contributing with brilliant ideas and support, making understanding a lot easier.

Contents

List of Abbreviations

BF	**B**eam**f**orming
BSS	**B**lind **S**ignal **S**eparation
DFT	**D**iscrete **F**ourier **T**ransform
DOA	**D**irection **O**f **A**rrival
DS	**D**elay and **S**um
DSP	**D**igital **S**ignal **P**rocessor
ITU	**I**nternational **T**elecommuncation **U**nion
MVDR	**M**inimum **V**ariance **D**istortionless **R**esponse
PESQ	**P**erceptual **E**valuation of **S**peech **Q**uality
PTZ	**P**an **T**ilt **Z**oom
RMVB	**R**obust **M**inimal **V**ariance **B**eamformer
SOI	**S**ignal **O**f **I**nterest
STOI	**S**hort-**t**ime **O**jective **I**ntelligibility
TDOA	**T**ime **D**ifference **O**f **A**rrival

List of Symbols

Symbol	Name	Unit
F_s	sample frequency	Hz
λ	wave length	m
θ	azimuth angle	rad/deg
ϕ	elevation angle	rad/deg
$\mathbf{d}$	distance vector	m
$\boldsymbol{\tau}$	steering vector	
M	number of microphones in array	
N	number of samples	
(t)	continuous time	
$[n]$	discrete time	

Chapter 1

Introduction

In this chapter, relevant background for this thesis is presented. Furthermore, a general description regarding digital signal processing is given. Finally, the outline and goal of this thesis are presented.

1.1 Background

Humans are on a daily basis exposed to sound sources of various types and from different directions. Since we have two ears we have the ability to tell the location of the sound source. We also have the ability, without much effort, to differentiate between intermixed speech sources, a phenomena called the Cocktail Party Effect which was defined and named by Colin Cherry in 1953 [1]. Imagine being at a cocktail party with some background music and a lot of chatter around you. Although our ears pick up every sound source the brain still manages to adjust easily when participating in a conversation with someone while, to certain degree, ignoring the rest.

Doing this digitally is a complex task that requires a lot of computations. For example, having a conversation over the phone is especially challenging when a person's speech signal is combined with some background noise. The signal of interest will be intermixed and more difficult to understand. Not being in the same room will leave the brain with less information to process and therefore a harder task to separate the sources.

A solution for this is to use a beamformer that attenuates the surrounding signals leaving the desired source signal untouched. This thesis will address the problem of suppressing noise by using a beamformer.

1.2 Problem Description

The following problem description is directly taken from the idea presented by Axis Communications AB:

> *"Axis Communications AB would like to investigate the feasibility for using a microphone array for far-field voice capture. Use cases include directing a PTZ-camera to a certain location and getting audio for that location using Beamforming techniques.*
>
> *We would also like to investigate the possibility of using Blind Signal Separation for noise cancellation purposes i.e. extract channels containing voice information and suppressing noise. A likely setup for prototyping will be a Raspberry Pi and a commercially available microphone array."*

1.2.1 Objective

As stated in the problem description this thesis attempts to address the problem of suppressing unwanted noise using beamforming techniques utilizing a microphone array. The objective is to implement and discuss different DOA and BF techniques and elucidate the underlying methods and algorithms.

1.3 Microphone Array

A microphone array is a set of microphones operating in tandem. Different spatial arrangements are possible when having more than one microphone i.e. linear, planar or three-dimensional (with the two latter requiring more than two and three microphones respectively). With a linear setup all the microphones are placed in a straight line in one dimension. Planar arrays have microphones distributed over two dimensions and lastly microphones can be placed freely in a three-dimensional array.

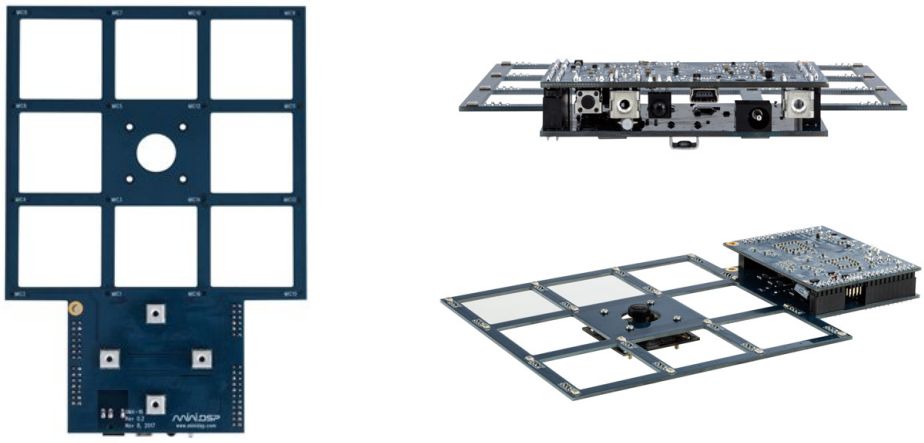


FIGURE 1.1: Microphone array used during the thesis.

The microphone array used during the thesis project is a planar 16 channel USB microphone array with plug & play USB audio connectivity [2] which can be seen in Figure 1.1. The array can also be used in a linear way by selecting a subset of microphones. Specifications for the microphone array used can be found in Appendix A.

1.4 Digital Signal Processing

Digital signal processing can be divided into three blocks as seen in Figure 1.2. This thesis is subject to the middle block i.e. the digital signal processor (DSP). All form of processing and filtering is programmed in python[1].

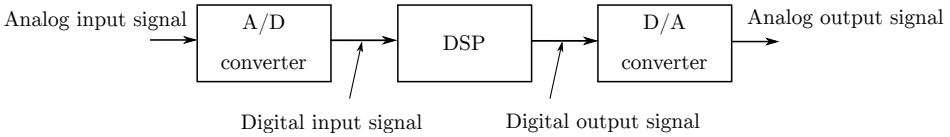


FIGURE 1.2: Block diagram of digital signal processing system.

1.5 Report Outline

This thesis is done in collaboration with Axis Communications AB. Axis is the market leader in network video [4] and focuses mainly on network solutions regarding surveillance and security.

The report contains signal processing techniques investigated during the thesis i.e. direction of arrival (DOA), beamforming (BF) and some future work ideas regarding blind signal separation (BSS). Chapter 2 contains a general introduction presenting relevant theory regarding the processing techniques at hand. Chapter 3 describes in detail different implementations of DOA and BF. The test setup is explained in Chapter 4 and it also covers the evaluation methods and units of measurements used during the thesis. The test results are presented in Chapter 5 and conclusions are drawn in Chapter 6. This chapter also contains proposals for future work.

1.5.1 Scope and Limitations

The algorithms will be evaluated using the commercially available microphone array described in Section 1.3 and various tests will be performed using an anechoic chamber and an ordinary room.

Due to the vast research area some limitations are introduced. Firstly, only one signal of interest (SOI) will be considered at a time. As the title suggests speech signal is prioritized and therefore the assumption is made that the SOI is not rapidly moving and the DOA does not vary significantly for small periods of time. Furthermore, the frequencies a voice signal is able to produce is roughly within range 250 to 6000 Hz where vowels are in the lower range and consonants in the higher range of the spectra [5]. Since the thesis focuses on far-field voice capture, we consider, in accordance with the telecommunications industry, the frequency ranges between 300 and 3400 Hz most important [6]. Both latter limitations will significantly decrease the amount of computations needed during the process.

[1] Programming Language [3]

1.6 Previous Work

Sensor array processing, including beamforming has been applied and share commonalities in a vast number of fields like radio, sonar, radar, wireless communications and acoustics. A great deal of work has gone into studying and developing beamforming techniques. They can be categorized into fixed (non-adaptive) and adaptive beamformers. In fixed beamforming techniques, the array geometry and weights are fixed in beforehand. The *delay-and-sum* technique is a fixed beamformer. In adaptive methods the algorithm adapts to the situation at hand. Common examples of adaptive beamformers is the *minimum variance distortionless response filter* (MVDR) and the *linearly constrained minimum variance filter* (LCMV) [7].

As to the closely related field of *blind signal separation* (BSS) there also exists much research. Statistical methods e.g. independent component analysis (ICA) are classically used to address the problem where sources are assumed to be uncorrelated. Furthermore, dependent component analysis (DCA) aim to extract source signals that are correlated. This can, for instance, be the case in multiple sensor systems where one can expect correlation in signals to adjacent sensors [8]. More recent work also includes methods that combine classical beamforming- and noise reduction-techniques with the uprising popular and effective methods of machine learning and neural networks [9].

Chapter 2

Theory

In this chapter basic concepts are thoroughly explained for further reference. Furthermore, relevant theory for direction of arrival (DOA) is presented. Different strategies and derivation of mathematical equations are expressed. Secondly, a general description and introduction into beamforming (BF) is given.

2.1 Introduction

Sound can be seen as a wavelike motion in some sort of elastic media such as air. It can also be seen as stimuli of the hearing mechanism by vibrating air particles [10]. Sounds can be described by the terms frequency and intensity, more commonly known as pitch and loudness, respectively. Humans are susceptible to frequencies between roughly 20 Hz to 20 kHz and intensities between 0 and 140 dB [5].

Sound can either travel in a direct path from the source to the receiver or be reflected via a surface on its way. When a sound is produced it will still be audible a short time after due to the reflections causing the signal to reach the receiver at different times. This phenomena is called reverberation. The time it takes for the signal to reach the receiver depends on the characteristics of the room. Absorbing materials will cause a fast decay and reflective surfaces will instead have a slow decay and the signal will linger longer [10].

2.1.1 Near-field vs. Far-field

A sound source is considered to emit a spherical wavefront. However, when the distance between source and receiver is far enough the incident wavefront can be considered planar as shown in Figure 2.1.

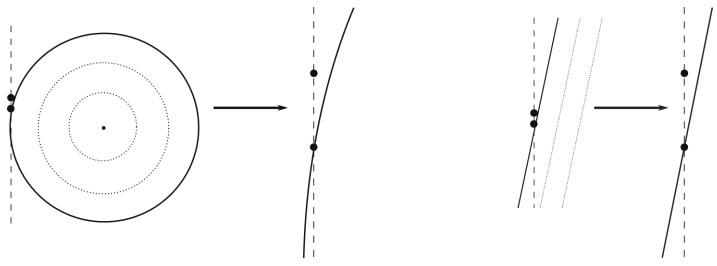


FIGURE 2.1: Wave fronts approaching 2 microphones in near-field (left) and far-field (right).

The transition between near- and far-field is dependent on the wavelength of the signal. If the distance r is strictly larger than the following expression the signal is considered to be far-field [11].

$$r > 2\lambda \tag{2.1}$$

For a source with multiple frequency components and thus multiple wavelengths the shortest wavelength is to be considered. The blue region in Figure 2.2 illustrates far-field for all frequencies.

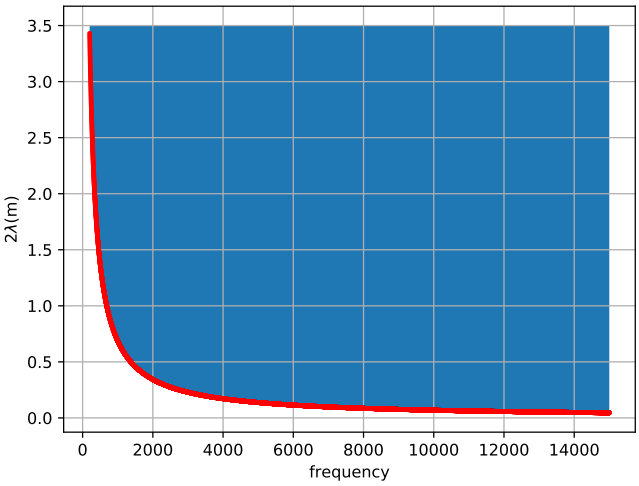


FIGURE 2.2: Far-field (blue region) and near-field (area under curve).

This work will only consider far-field regions and thus the wavefront is assumed to be planar hereafter.

Signal Representation and Sampling

Signals can be represented in continuous and discrete time respectively.

$$x(t), \quad t \in \mathbb{R} \quad \longleftrightarrow \quad x[n] = x(nT_s), \quad n \in \mathbb{Z} \tag{2.2}$$

Where T_s is the sampling period which needs to fulfill the Nyquist Sampling theorem [12].

$$\frac{1}{T_s} = F_s > 2 \cdot F_{max} \tag{2.3}$$

Where F_s is the sampling frequency and F_{max} is the highest frequency present in the analog signal. Failing to fulfill the requirement could result in aliasing, a form of distortion. Higher frequencies will not be correctly recomposed and instead be presented as false frequency components in the outgoing signal.

2.1.2 Frequency Domain

It is often practical to express a signal in terms of its frequency components. Signals are transferred via discrete fourier transform (DFT) into frequency-domain which can be written as

$$X[k] = \sum_{n=0}^{N-1} x[n] e^{\frac{-j2\pi nk}{N}} \tag{2.4}$$

For some applications calculations in the frequency-domain is simpler than the corresponding calculation in time-domain. For example as can be seen in Table 2.1 a time shift (delay) and convolution[2] are simple multiplication operation in the frequency-domain and will therefore reduce computational complexity.

Continuous time		**Frequency**
$x(t-\tau)$	$\longleftrightarrow$	$X(\omega) \cdot e^{-j2\pi\omega\tau}$
$x(t) * y(t)$	$\longleftrightarrow$	$X(\omega) \cdot Y(\omega)$

TABLE 2.1: Relations between time-domain and frequency-domain.

2.1.3 Delays

Delaying a signal in time-domain corresponds to phase shifting in frequency-domain. The first relation in Table 2.1 can be written in discrete time form as

$$x[n-\tau] \quad \longleftrightarrow \quad e^{\frac{-j2\pi k\tau}{N}} X[k] \tag{2.5}$$

Fractional Delay

To be able to steer the beam in the desired direction it is needed to delay the incoming signals of the microphone array to align the wavefront of interest. However, since the signal is sampled and stored as a discrete signal the minimum delay is limited to the sample period T_s. This will result in limited accuracy since the delays only can be a multiple of whole samples.

To achieve higher accuracy, any fractional part of the delay must be accounted for which can be done by reconstructing the signal using fractional delay [13]. The idea behind fractional delay is to shift the integer part of the delay the correct amount of samples. The fractional part is then created by resampling the signal.

2.1.4 Microphone Array

The definition of the coordinates system used in this thesis is given in Figure 2.3. Azimuth, $\{\theta \in \mathbb{R} | 0 \leq \theta < 2\pi\}$, is defined as the angle on the xy-plane starting from the x-axis and increasing in the counterclockwise direction. Elevation, $\{\phi \in \mathbb{R} | 0 \leq \phi \leq \frac{\pi}{2}\}$, is defined as the angle between the z-axis and the xy-plane starting from the z-axis and increasing towards the xy-plane. The radius, r, is the length from the point of origin to the point of interest (r, θ, ϕ).

[2](*) denotes the convolution operator.

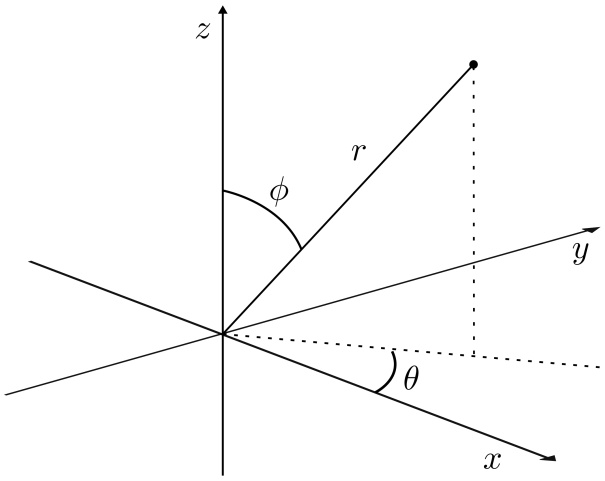


FIGURE 2.3: Coordinate system used.

For the array used throughout our thesis the orientation and enumeration can be seen in Figure 2.4.

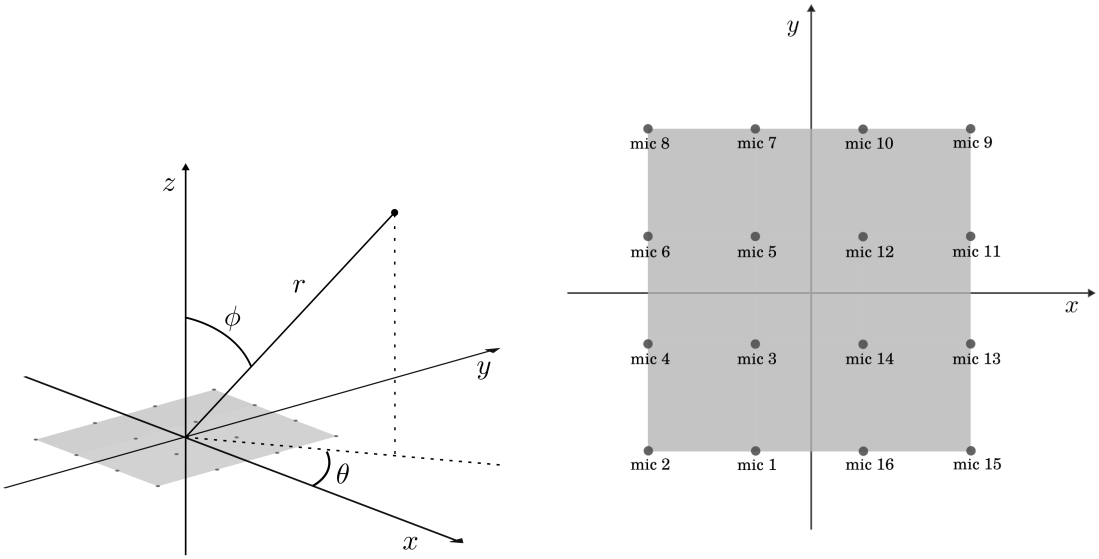


FIGURE 2.4: Orientation and enumeration of microphone array.

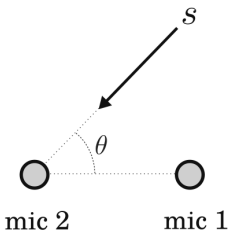


FIGURE 2.5: Simple microphone array and a propagating sound source.

In Figure 2.5 a simple microphone array consisting of two microphones is illustrated. A sound source s is propagating towards the microphone at an angle and is expected to arrive at the microphones at different times. This time difference is referred to as the time difference of arrival (TDOA) and plays an essential part in the field of microphone array processing and beamforming.

We obtain the TDOA, denoted τ between microphone a and b as shown in the left Figure 2.6 by calculating

$$\tau = \frac{d}{c} = \frac{l \cdot cos\theta}{c} \tag{2.6}$$

where d is the travel difference, c is the speed of sound[3], θ is the incident angle of the sound wave, and l the distance between the two microphones.

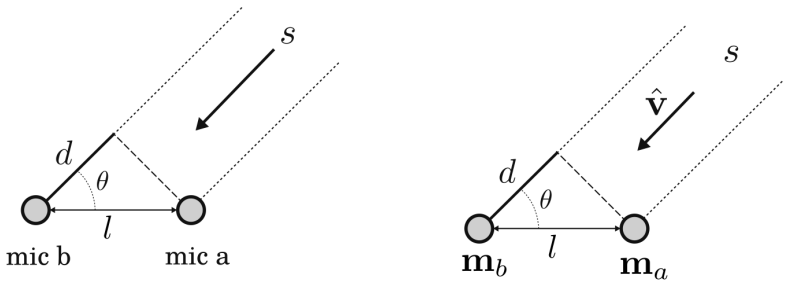


FIGURE 2.6: Different representations of a simple microphone array and a propagating source.

Given that we know the direction of the source, d in Figure 2.6 can be calculated in another way by using the unit vector $\hat{\mathbf{v}} = \begin{pmatrix}\sin\theta & \cos\theta\end{pmatrix}$ that has the same direction as the source. Further more, we denote the spatial positions of the microphones in Figure 2.6 as

$$\mathbf{m}_a = \begin{pmatrix} x_a \\ y_a \end{pmatrix}, \quad \mathbf{m}_b = \begin{pmatrix} x_b \\ y_b \end{pmatrix} \tag{2.7}$$

[3]The speed of sound in air at 20°C is 343 m/s.

We can then calculate d using

$$d = \begin{pmatrix} \cos\theta & \sin\theta \end{pmatrix} \begin{pmatrix} x_a - x_b \\ y_a - y_b \end{pmatrix} \tag{2.8}$$

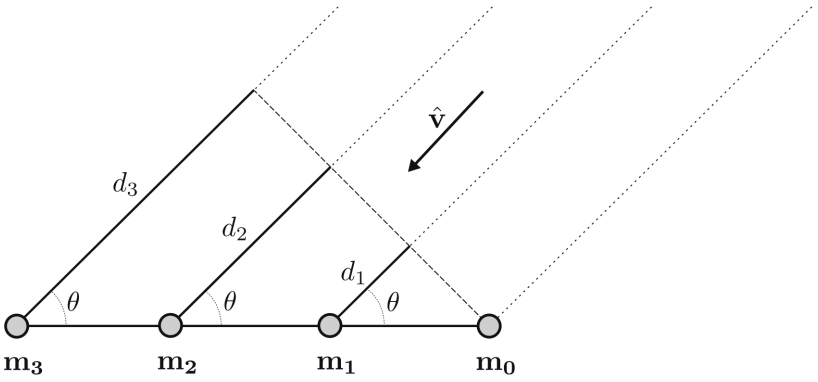


FIGURE 2.7: Microphone array - positions and propagation.

Figure 2.7 shows a situation with 4 microphones where $\mathbf{m}_0$ is the spatial position of what we call the reference microphone. This means that d_1, d_2, d_3 will be calculated in relation to this microphone in the same way as in equation 2.6.

So far, only two dimensions have been considered. As this thesis deals with the three dimensional case, the following shows how we can determine the distances and TDOA for an array and the direction of a propagating sound wave in 3D. The direction of the propagating wave can be expressed as the unit vector

$$\hat{\mathbf{v}}(\theta, \phi) = \begin{pmatrix} \cos(\theta)\sin(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\phi) \end{pmatrix}^T \tag{2.9}$$

where θ and ϕ is the azimuth and elevation angles respectively, as illustrated in Figure 2.8.

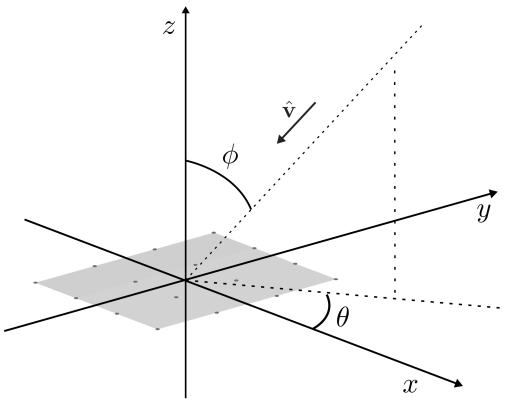


FIGURE 2.8: Direction of propagating wave in relation to microphone array orientation.

An array geometry consisting of M microphones and a reference microphone located in $\begin{pmatrix} x_{ref} & y_{ref} & z_{ref} \end{pmatrix}^T$ is represented as

$$\mathbf{R} = \begin{pmatrix} x_{ref} - x_0 & x_{ref} - x_1 & \dots & x_{ref} - x_{M-1} \\ y_{ref} - y_0 & y_{ref} - y_1 & \dots & y_{ref} - y_{M-1} \\ z_{ref} - z_0 & z_{ref} - z_1 & \dots & z_{ref} - z_{M-1} \end{pmatrix} \tag{2.10}$$

where each microphone's spatial position is given by one column. We can compute the distances by projecting each microphone position onto the distance vector as

$$\begin{aligned} \mathbf{d} &= \hat{\mathbf{v}}\mathbf{R} \\ &= \begin{pmatrix} \cos(\theta)\sin(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\phi) \end{pmatrix}^T \begin{pmatrix} x_{ref} - x_0 & x_{ref} - x_1 & \dots & x_{ref} - x_{M-1} \\ y_{ref} - y_0 & y_{ref} - y_1 & \dots & y_{ref} - y_{M-1} \\ z_{ref} - z_0 & z_{ref} - z_1 & \dots & z_{ref} - z_{M-1} \end{pmatrix} \\ &= \begin{pmatrix} d_0 & d_1 & \dots & d_{M-1} \end{pmatrix} \end{aligned} \tag{2.11}$$

And to get the time delays we just divide by the speed of sound c and obtain

$$\frac{\mathbf{d}}{c} = \begin{pmatrix} \tau_0 & \tau_1 & \dots & \tau_{M-1} \end{pmatrix} = \boldsymbol{\tau} \tag{2.12}$$

Note that $\mathbf{d}$, and therefore $\boldsymbol{\tau}$ may well be containing negative values. We interpret this as the time of arrival being earlier than at the reference microphone. $\boldsymbol{\tau}$ if often called the steering vector.

2.1.5 Signal Model

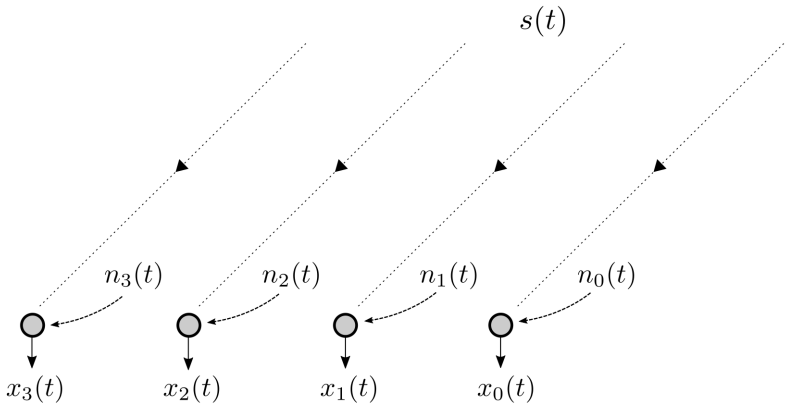


FIGURE 2.9: Anechoic signal model under far-field assumption.

The input signal at time t received at the i-th microphone in the array can be modeled as

$$x_i(t) = \alpha_i s(t - \tau_i) + n_i(t) \tag{2.13}$$

where α_i is an attenuation factor, $s(t)$ denotes the signal of interest with τ_i being the relative delay to the reference microphone. It can be noted that τ is always zero for the reference microphone. n_i denotes any noise in the environment and from the microphone itself. The output $y(t)$ of a microphone array with M microphones can simply be seen as the sum of the individual signals, here normalized with M

$$y(t) = \frac{1}{M} \sum_{i=0}^{M-1} x_i(t) \tag{2.14}$$

We can also express the time-domain model in equation 2.13 by transforming it to frequency-domain as [7]

$$X_i(\omega) = \alpha_i(\omega) e^{-j\omega(\tau_i)} S(\omega) + N(\omega) \tag{2.15}$$

and let

$$\mathbf{X}(\omega) = \begin{bmatrix} X_0(\omega) & X_1(\omega) & \dots & X_{M-1}(\omega) \end{bmatrix}^T \tag{2.16}$$

The signal model we have presented in equation 2.13 and in Figure 2.9 is referred to as an anechoic signal propagation model. This means that in addition to environmental noise, we only take the signal attenuation and propagation delay into account. Most commonly however, there are more effects in play. Depending on the surroundings, effects such as reverberation and multi-path propagation (reflections) may be present. To account for this, another, more comprehensive representation can be used

$$x_i(t) = w_i(t) * s(t) + n_i(t) \tag{2.17}$$

Here the microphone output $x(t)$ is represented as a convolution of the source signal $s(t)$ and the filter $w(t)$ which encompasses the effects described above. The meaning of the noise term $n(t)$ is kept equivalent to the one in equation 2.13 [14].

In this work, the anechoic signal model will be used primarily to describe and approach methods mathematically.

2.2 Direction of Arrival

Within signal processing direction of arrival (DOA) denotes from which direction a propagating wave is arriving at a certain point of incidence. If the propagating wave is caused by a source that if far away (far-field) the wave is considered to be planar. Together with the array these will form an angle of incidence which can be uniquely defined as shown in Figure 2.10 with θ as the angle of incidence.

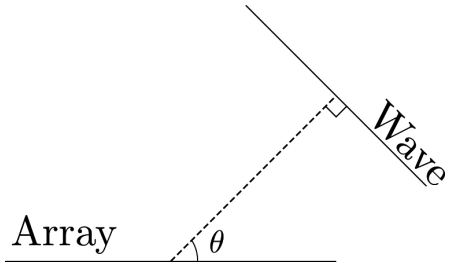


FIGURE 2.10: Angle of incident.

This works for two dimensions. The DOA for a three dimensional space is described by adding another angle as can be seen in Figure 2.3. Consider (r, θ, ϕ) to be a far-field source and the origin to be the point of incident.

2.2.1 Determining DOA

To find a source direction of arrival we rely on estimating the TDOA $\boldsymbol{\tau}$ for each microphone in the array (see equations 2.6, 2.12). There are many methods for obtaining TDOA [7]. In this thesis we make use of cross correlation which will be described in more detail later on.

Assuming we know the TDOA for each microphone, along with its spatial coordinates, we can try and calculate the DOA. However, the geometry and number of microphones determines the possible directions that the source propagates from. To illustrate this, consider the two-dimensional configuration in Figure 2.11.

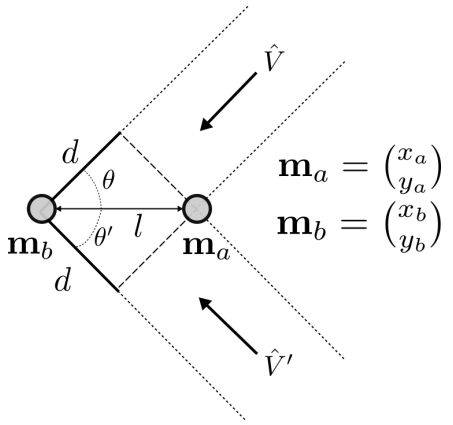


FIGURE 2.11: Possible angles of incident.

Here, both θ and θ' are possible angles for calculating d i.e.

$$d = \begin{pmatrix} \cos\theta & \sin\theta \end{pmatrix} \begin{pmatrix} x_a - x_b \\ y_a - y_b \end{pmatrix} = \begin{pmatrix} \cos\theta' & \sin\theta' \end{pmatrix} \begin{pmatrix} x_a - x_b \\ y_a - y_b \end{pmatrix} \tag{2.18}$$

In order to find an unique solution for d, the array must contain at least three microphones in which their coordinates are non-collinear, e.g an L-shaped array. Similarly, for three dimensions, the array must contain at least four microphones located not in the same plane to uniquely determine θ, ϕ. In this thesis we consider the three dimensional case but as mentioned before, we have a planar array. Consequently,

we cannot determine DOA uniquely, but rather as two possible solutions. Using equation 2.11 to find θ, ϕ we can for instance solve the system

$$\begin{cases} \cos(\theta)\sin(\phi)(x_{ref} - x_1) + \sin(\theta)\sin(\phi)(y_{ref} - y_1) + \cos(\phi)(z_{ref} - z_1) = d_1 \\ \cos(\theta)\sin(\phi)(x_{ref} - x_2) + \sin(\theta)\sin(\phi)(y_{ref} - y_2) + \cos(\phi)(z_{ref} - z_2) = d_2 \end{cases} \tag{2.19}$$

As we cannot know the source position (far-field), we choose to narrow our possible solutions to be located on distance $r = 1$ away from the origin i.e on the surface of the unit sphere. Each row in the system above then has infinite solutions for θ, ϕ, all located on a circle on the unit-sphere. We will refer to this as a *candidate circle* and solving the above system translates in to finding intersections of two such circles as illustrated in Figure 2.12.

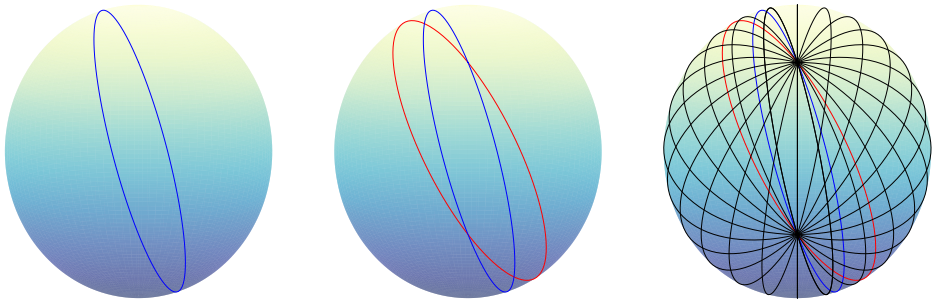


FIGURE 2.12: Unit sphere with possible solutions for θ, ϕ along the blue circle (left) and the intersections between the blue and red circle (middle). The rightmost sphere shows intersections between 15 circles.

The same procedure holds for an array with a larger number of microphones. The rightmost sphere in Figure 2.12 illustrates the candidate circles for a 16-mic array. Note that in theory, having more candidate circles than two, does not provide any additional information. However in the applied scenario, we will see that more microphones can be of help to increase the accuracy of the result.

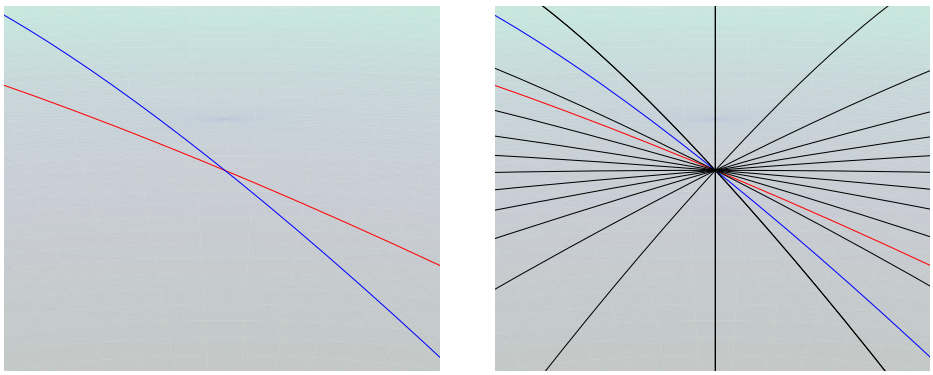


FIGURE 2.13: Sphere with 2 and 15 intersecting lines zoomed in respectively.

2.2.2 Cross Correlation

Cross correlation is a measure of similarity between two signals [12]. If the distance between the source and a microphone, a, differs from the distance between the source and another microphone, b, the signal will reach the microphones at different times. Some noise will also be added to each of the arriving signals. However, since the source signal is correlated and the noise is assumed not to be, we can use cross correlation for the signals.

$$r_{a,b}(\tau) = E[x_a(t) \cdot x_b(t-\tau)] \tag{2.20}$$

The TDOA can now be estimated with help from cross correlation by

$$\hat{\tau}_{a,b} = \arg\max_{\tau} r_{a,b}(\tau) \tag{2.21}$$

Where $\hat{\tau}_{a,b}$ is the estimated TDOA between the signals from microphone a and b. Since the signals are sampled before being cross correlated the precision of the correlation will be limited to the sample frequency. An increased sampling rate will result in a higher precision of the TDOA.

2.3 Beamforming

We can formulate the goal of a beamformer in words as capturing as much as possible from the SOI, meanwhile suppressing noise and unwanted sources. Denoting the beamformer output as y_{bf} and using the array signal model in equation 2.13 we can express the goal as constructing a filter $w_i(t)$ such that

$$y_{bf}(t) \approx s(t) \quad \text{where} \quad y_{bf}(t) = \frac{1}{M} \sum_{i=0}^{M-1} x_i(t) * w_i(t) \tag{2.22}$$

Further more, given the model in frequency-domain (equation 2.15), and with a chosen weight vector $\mathbf{W}$

$$\mathbf{W}(\omega) = \begin{bmatrix} W_0(\omega) & W_1(\omega) & \dots & W_{M-1}(\omega) \end{bmatrix}^T$$

we can compute the output of the beamformer at frequency ω to be

$$\mathbf{Y}_{bf}(\omega) = \mathbf{W}^T(\omega)\mathbf{X}(\omega) \tag{2.23}$$

Then the corresponding problem becomes to select the weights $\mathbf{W}(\omega)$ such that the output of the beamformer is a good approximation of the SOI, that is

$$\mathbf{Y}_{bf}(\omega) \approx S(\omega)$$

2.3.1 Narrow-band Beamforming

For narrow-band beamforming only a small portion or one frequency is considered. As can be seen in Figure 2.14 spacing between microphones matter when choosing a setup for BF purposes. With larger spacing the main lobe will become narrower and an increasing amount of side lobes will be present.

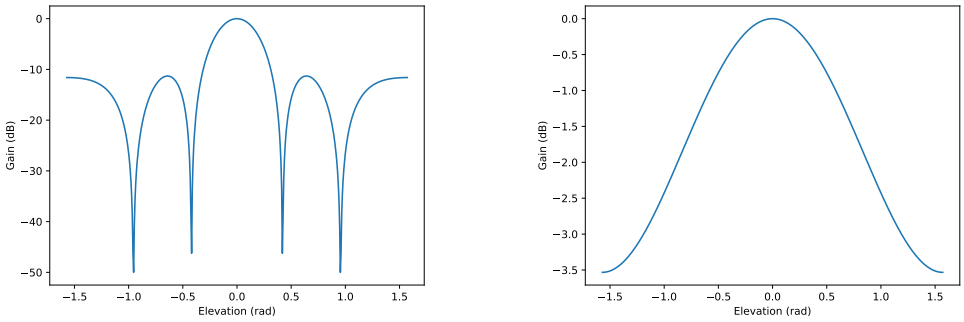


FIGURE 2.14: Frequency response with a frequency of 1000Hz for 16 microphones confined to a matrix, spaced with 21cm and 4.2cm respectively.

For a higher frequency more lobes with the same amplitude are present as can be seen i Figure 2.15.

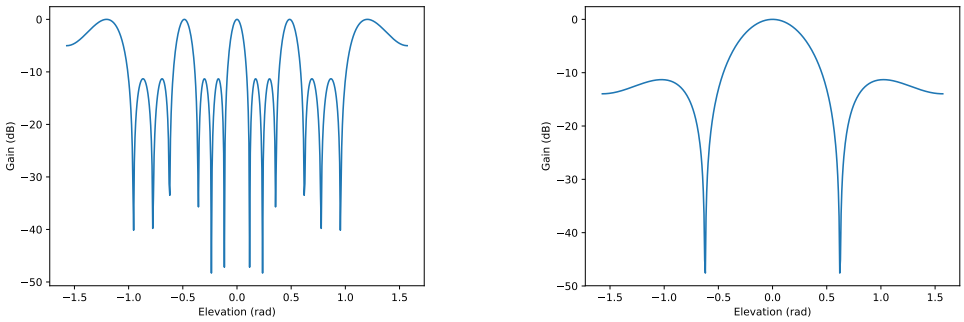


FIGURE 2.15: Frequency response with a frequency of 3500 for 16 microphones confined to a matrix, spaced with 21 cm and 4.2 cm respectively.

This is called spatial aliasing and can also occur depending on the microphone array aperture. Since the wavelength of a high frequency signal is shorter it will be able to phase shift more than one wavelength on its way between two microphones. This means that the directionality of the incoming signal can not be determined. To avoid spatial aliasing the microphone spacing l needs to satisfy [15]

$$l \leq \frac{c}{2f} \tag{2.24}$$

With a spacing of $l = 0.042$ m, for our microphone array, this holds for frequencies under 4083 Hz. Since the interesting frequencies of speech only reaches about 3.4

kHz this will not be a problem in our case. The goal is to have a narrow main lobe without the presence of grating lobes.

Delay and Sum

Delay and sum is the most simple and oldest beamforming technique still used today [16]. The idea behind the delay-and-sum technique is to delay each microphone signal as if the SOI would have reached each microphone at the same time. By then adding the delayed signals together there is constructive interference in the source direction, and attenuation due to destructive interference in all other directions.

Assuming that we know the source direction of arrival and have obtained the delays τ_i we can construct a delay-and-sum beamformer as follows: Firstly we delay each signal x_i with $-\tau_i$, canceling out all of the relative delays. And secondly summing the signals to produce the beamformed output y_{DS}.

$$\begin{aligned} y_{DS}(t) &= \frac{1}{M}\sum_{i=0}^{M-1} x_i(t-(-\tau_i)) \\ &= \frac{1}{M}\sum_{i=0}^{M-1} \alpha_i s((t+\tau_i)-\tau_i) + n_i(t+\tau_i) \\ &= \frac{1}{M}\sum_{i=0}^{M-1} \alpha_i s(t) + n_i(t+\tau_i) \end{aligned} \tag{2.25}$$

The DS beamformer can also be performed in frequency-domain by using the relationship in equation 2.5 where we phase shift over all frequencies. Below it is shown for discrete frequencies, k, where τ_i is assumed to be a known integer sample delay.

$$Y_{DS}[k] = \frac{1}{M}\mathbf{W}^T[k]\mathbf{X}[k] \quad \text{where} \quad \mathbf{W}[k] = \begin{pmatrix} e^{\frac{j2\pi k\tau_0}{N}} & e^{\frac{j2\pi k\tau_1}{N}} & \dots & e^{\frac{j2\pi k\tau_{M-1}}{N}} \end{pmatrix}^T \tag{2.26}$$

Minimum Variance Distortionless Response

The Minimum Variance Distortionless Response (MVDR) beamformer, commonly referred to as Capon's method [17], is one of the most often used adaptive beamforming methods [7]. The idea is to choose a set of weights, $\boldsymbol{W}(\omega)$, such that the output power is minimized whilst the gain in direction of the signal is prevented from being reduced. Mathematically the problem of choosing a set of weights can be written as[4]

$$\underset{\boldsymbol{W}(\omega)}{\text{minimize}}\, \boldsymbol{W}^H(\omega)\boldsymbol{R}_{XX}\boldsymbol{W}(\omega) \quad \text{s.t.} \quad \boldsymbol{W}^H(\omega)\boldsymbol{a}(\theta,\phi) = 1 \tag{2.27}$$

[4]where ^{H} denotes the complex conjugate also known as the Hermitian transpose

which with help of Lagrange multipliers can be shown to have the analytical solution [17] given by

$$\boldsymbol{W}(\omega) = \frac{\boldsymbol{R}_{XX}^{-1}\boldsymbol{\alpha}(\theta,\phi)}{\boldsymbol{\alpha}^{H}(\theta,\phi)\boldsymbol{R}_{XX}^{-1}\boldsymbol{\alpha}(\theta,\phi)} \tag{2.28}$$

$\boldsymbol{R}_{XX}$ is the power spectral density matrix for the incoming signals at each microphone, $\boldsymbol{\alpha}(\theta,\phi)$ contains information regarding direction and attenuation for each channel respectively. The output at a given frequency k can now be expressed as

$$Y_{MVDR}[k] = \boldsymbol{W}^{H}[k]\boldsymbol{X}[k] = \frac{\boldsymbol{\alpha}^{H}(\theta,\phi)\boldsymbol{R}_{XX}^{-1}\boldsymbol{X}[k]}{\boldsymbol{\alpha}^{H}(\theta,\phi)\boldsymbol{R}_{XX}^{-1}\boldsymbol{\alpha}(\theta,\phi)} \tag{2.29}$$

Where $\boldsymbol{X}[k]$ is the vector containing incoming signals for each microphone at a given frequency k. Transferring back into time-domain yields $y_{MVDR}(t)$.

Robust Minimum Variance Beamformer

Robust Minimum Variance Beamformer (RMVB) is an extension to MVDR. Instead of focusing in a particular direction the algorithm will focus on a set of specified directions. This will make the area of arrival larger and the signal will not be attenuated if the steering vector is of by a couple of degrees. To achieve this, the design is not limited to the direction of the steering vector but instead to minimize the uncertainty set which is formed by an ellipsoid that covers the number of possible directions of the steering vector [18].

2.3.2 Wide-band Beamforming

Different form narrow-band beamforming all the frequencies in a signal must be taken into account for wide-band beamforming. Since different frequencies have different wavelengths and will experience either negative or positive interference whilst being summed depending on the direction of arrival. To illustrate this simulated frequency responses are plotted for various microphone array setups and angles of incidence.

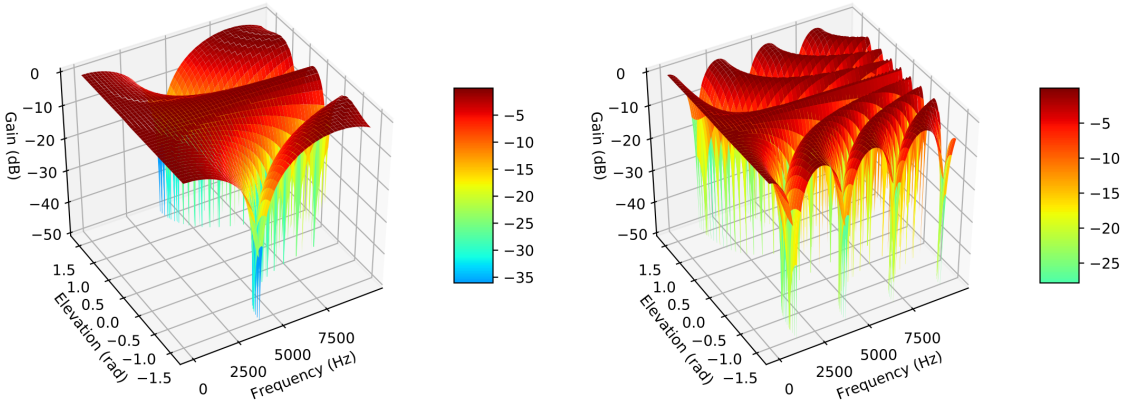


FIGURE 2.16: Frequency response for 4 microphones confined to a square, spaced with 4.2 cm and 12.6 cm respectively.

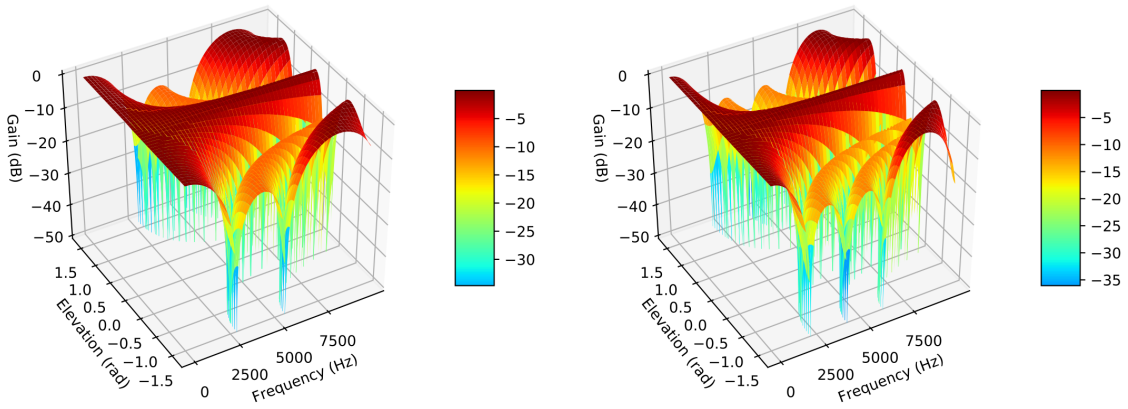


FIGURE 2.17: Frequency response for 9 and 16 microphones respectively confined to a square (matrix), spaced with 4.2 cm.

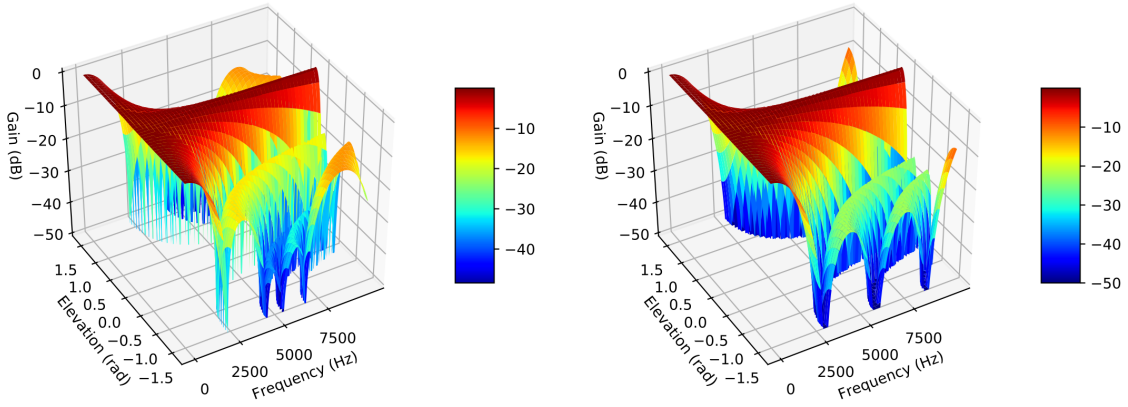


FIGURE 2.18: Frequency response for 16 microphones, spaced with 4.2 cm, with azimuth angle at 22.5° and 45° respectively.

The microphone array used during the thesis has got an equal spacing of 4.2 cm between its 16 microphones and the frequency response can be seen in Figure 2.17. Since the use case is limited to voice capture which contains frequencies between 300 and 3400 Hz, the side lobes are suppressed with a factor of at least -10 dB and will therefore only have less influence on the output. A more detailed frequency response for 1000 Hz and 3500 Hz can be seen in the right sub-figure of Figure 2.14 and 2.15 respectively.

Subband Beamforming

For wide-band beamforming to be possible the frequency response has to be uniform. As can be seen in the frequency response plots it is non-uniform which means that different frequencies will be attenuated differently and will present as strange artifacts in the output signal. Therefore, a response-invariant beamformer with a constant beamwidth has to be created to be able to handle broadband information [7].

One way to achieve a constant beamwidth is to divide the frequency spectra into subbands so that for a certain portion the beamwidth can be estimated as constant, so called narrow-band decomposition. The wideband signals of each channel, $X_i(\omega)$

are decomposed into L narrow-band (superscripted) signals so that each of the bands only contain a narrow range of frequencies. This is illustrated in Figure 2.19. For each subband, l, all contributions from each of the M channels respectively are added. After processing each subband the signal is then combined into a wide-band output signal, $Y_{bf}(\omega)$.

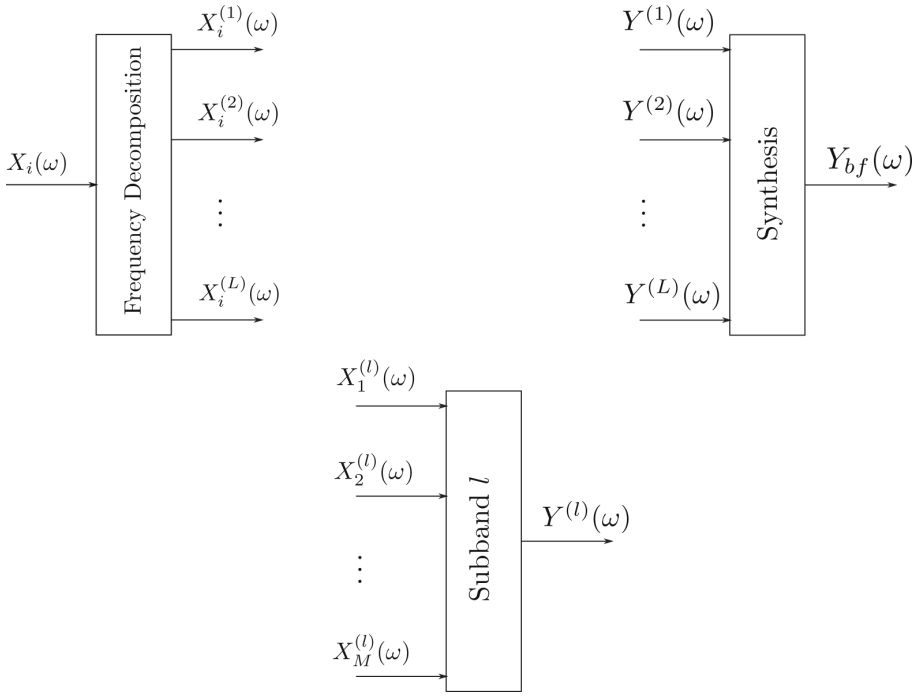


FIGURE 2.19: Frequency decomposition.

Chapter 3

Method

This chapter contains details regarding the implementation of different methods used in this thesis. The programming language python has been used to implement, evaluate and test our work.

3.1 Direction of Arrival

An algorithm for estimating direction of arrival in real time has been implemented. As described in section 2.2 we make use of cross correlation as well as finding intersections between a set of candidate circles. Due to its real-time purpose, the algorithm contains a pre-computation step and an online-computation step. The goal of the pre-computation step is to minimize necessary calculations in real-time. We can summarize the algorithm on a high level in words:

Pre computation:
1. Create and compute a map between candidate DOAs and distances.

Online computation:
1. Obtain $\mathbf{d}'$ from cross correlation.
2. Look up the possible candidates (circles) for each $d'_k \in \mathbf{d}'$.
3. Find intersection points on half-sphere.
4. Extract DOA as θ, ϕ from found points.

We continue below to describe the steps in a bit more detail.

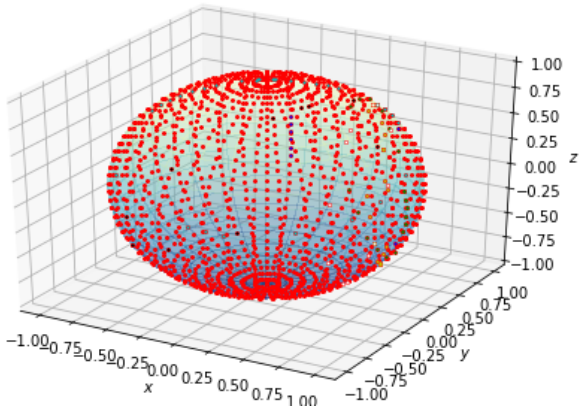


FIGURE 3.1: 40x40 points, evenly distributed on a unit sphere.

Build sphere

As illustrated in Figure 3.1 above, we compute and store a representation of a sphere consisting of $res_\theta \cdot res_\phi$ points evenly distributed on the surface of a unit sphere.

$$\sum_\theta \sum_\phi \hat{v}(\theta, \phi), \qquad \begin{cases} \theta = 0, \alpha, 2\alpha, \ldots, \text{res}_\theta \cdot \alpha & \alpha = \frac{2\pi}{\text{res}_\theta} \\ \phi = 0, \beta, 2\beta, \ldots, \text{res}_\phi \cdot \beta & \beta = \frac{\pi}{\text{res}_\phi} \end{cases} \tag{3.1}$$

Then for all points, calculate and store $d_k, \hat{\mathbf{v}}(\theta, \phi)$, where $d_k \in \mathbf{d}$

$$\sum_\theta \sum_\phi \sum_{k=0}^{M-1} \hat{\mathbf{v}}(\theta, \phi) r_k \text{ or } \sum_\theta \sum_\phi \hat{\mathbf{v}}\mathbf{R} = \sum_\theta \sum_\phi \mathbf{d} \tag{3.2}$$

and $r_k = \begin{pmatrix} x_{ref} - x_k & y_{ref} - y_k & z_{ref} - z_k \end{pmatrix}^T$ is the column vector containing the k-th microphone's relative position (see equation 2.10).

Candidate circles

Now given an input d'_k in $\mathbf{d}'$ we can look up the set of stored $\hat{\mathbf{v}}(\theta, \phi)$ which constitutes a candidate circle for each d'_k. However, as the real input value is not likely to be exactly equal to any of the stored values we pick a d_k so that it fulfills $\mid d'_k + u \mid \in d_k$ for some deviation u. How we select u depends on the resolution of the sphere. The left sub-figure in Figure 3.2 shows an example of a candidate circle for a particular d_k.

Intersections - least squares

Now that we have our candidate circles, we go through the points pairwise and find where each consecutive pair intersects using least squares. The right sub-figure in Figure 3.2 shows the candidate circles of an input $\mathbf{d}'$ and their intersections. Note that in actuality we only care about the half sphere as to the planar geometry of our microphone array.

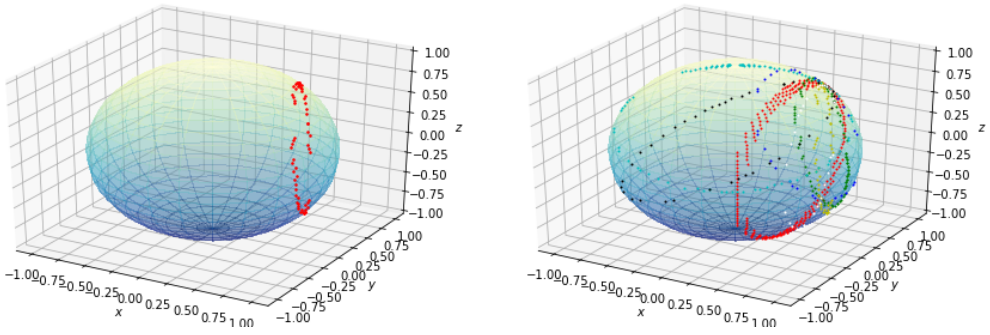


FIGURE 3.2: Points contained in a candidate circle (left) and multiple intersecting candidate circles (right).

3.2 Beamforming

Two beamformers have been implemented in python and are described in more detail below. Each of the methods are tested according to the setup in Chapter 4.

3.2.1 Delay and Sum

The DS beamformer has been implemented, one operating in time-domain (equation 2.25) and one in frequency-domain (equation 2.26). From a selected direction θ, ϕ, in which to direct the beam, the steering vector $\boldsymbol{\tau}$ (equation 2.12) is calculated. Each incoming channel is then delayed appropriately and summed to the produced output. Since the signal is discrete we can only allow a $\boldsymbol{\tau}$ which consists of delays that correspond to an exact number of samples i.e. integers. To overcome this we have made an implementation of fractional delays (section 2.1.3). Tests have been performed both with and without the use of fractional delays and the results are discussed in Chapter 5.

3.2.2 MVDR

One MVDR beamformer has been implemented. The MVDR-weights are added to filter out unwanted noise coming from other directions. To increase computational effectiveness the entire recording of 16 channels is delayed according to the steering vector $\boldsymbol{\tau}$ using fractional delay and transferred to the frequency-domain before applying the weights. Then, for each frequency k and each of the M channels weights are added as described in equation 2.29. This results in a filtered output which then is transferred back into time-domain. Test results of this method are found in Chapter 5.

Chapter 4

Testing

This chapter describes the testing procedure. Firstly, the test setup is explained and test cases and environment are described. Finally, units of measurements will be introduced to be able to compare the results.

4.1 Setup

Testing and evaluation is performed at two different locations. Firstly at Axis in a small studio room with dampening elements on the wall. The studio is shown in Figure 4.1. Secondly in an anechoic chamber at LTH seen in Figure 4.2. Both of the rooms have a fast decay from a reverberation perspective.

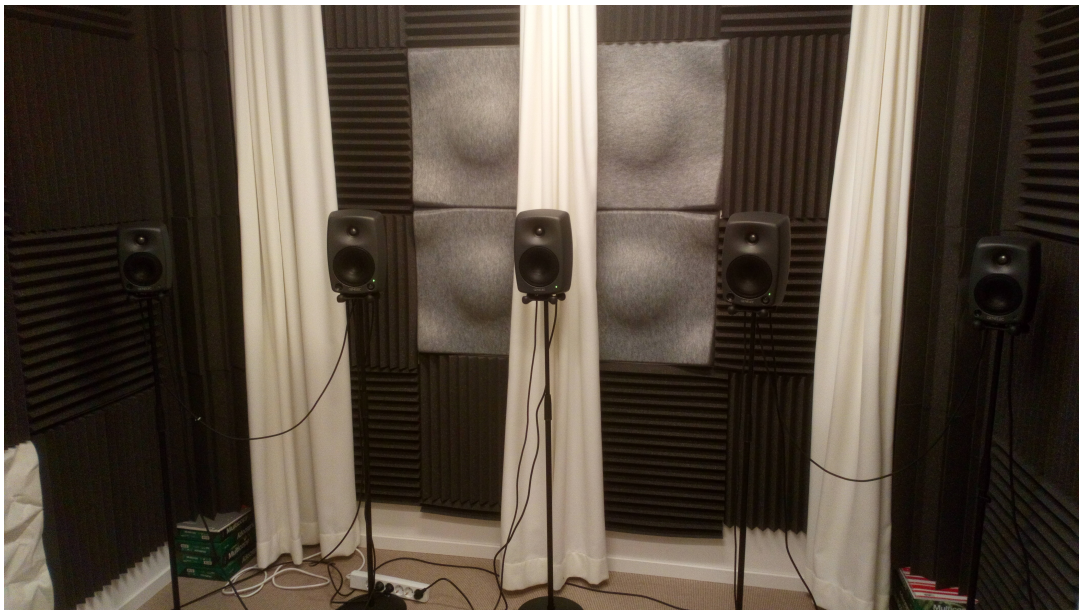


FIGURE 4.1: Demonstration of setup.

FIGURE 4.2: Demonstration of setup.

The test setup consists of five speakers located at a distance of 2.5 m from the microphone array. The speakers are placed in an arc of a circle with an angle of $\pm 15^\circ$ and $\pm 30^\circ$ from the main line, respectively. The height of the speaker and the array are the same. A sketch of the test setup is shown in Figure 4.3.

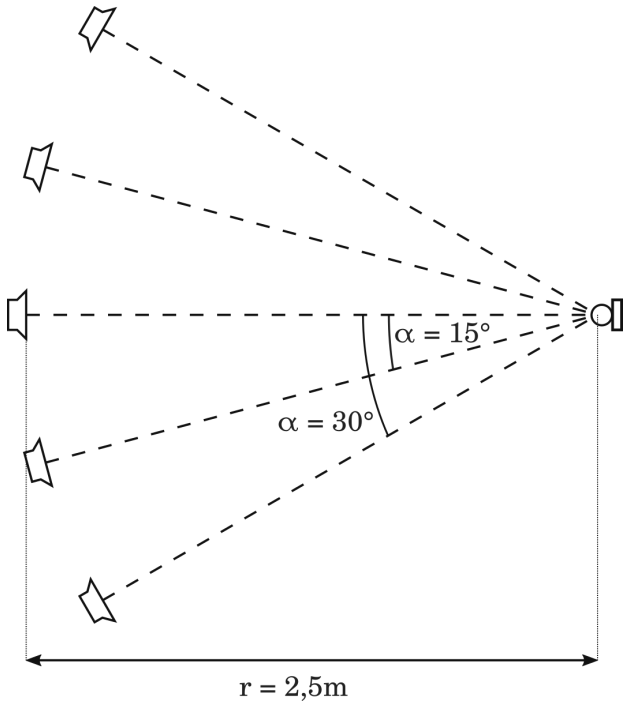


FIGURE 4.3: Demonstration of setup.

The SOI is always played through the speaker in the middle while the noise is coming from the four other speakers. A test sequence is recorded per permutation of test cases and saved for processing with different BF algorithms. The result are shown in the next chapter.

4.2 Test Cases

To limit the testing only a couple of directions are considered. The directions of interest are displayed in Table 4.1. During the tests in the studio three test sequences were recorded for each set of angles. The first sequence with noise only, second with SOI only and the last with SOI and noise combined. The SOI is a recording of four different Harvard sentences, two spoken by a male person, and two by a female as seen in Table 4.2. The Harvard sentences are a collection of phrases that are used for standardized testing different types of telephone systems [19]. The noise used in this case was white noise. For each sequence two different sample frequencies were used.

Azimuth, θ	**Elevation, ϕ**	**Sampling rate**	**Test sequence**
0°	0°	44.1 kHz	white noise
30°	30°	16 kHz	SOI
60°	60°		SOI + white noise
90°	90°		*sine tones*
			SOI + sine tones

TABLE 4.1: Test cases - each possible combination of the columns.

During testing in the anechoic chamber two test sequences were added for each set of angles these are displayed as italicized in the table. In addition to white noise a single sinus tone of 500 Hz, 1 kHz, 2 kHz and 3 kHz was used as noise. This gives a total of 96 test recordings for the studio and 160 for the anechoic chamber.

Sentence	**Speaker**
It's easy to tell the depth of a well.	Female
Kick the ball straight and follow through.	Male
Glue the sheet to the dark blue background.	Female
A pot of tea helps to pass the evening.	Male

TABLE 4.2: Harvard sentences.

4.3 Units of Measurement

In order to compare the results, units of measurement are used. Short-time Objective Intelligibility and Perceptual Evaluation of Speech Quality are used as well as subjective listening to any improvement after beamforming.

4.3.1 Short-time Objective Intelligibility

Short-time Objective Intelligibility (STOI) is an algorithm which calculates the speech intelligibility by comparing the processed signal with a reference signal. Instead of basing the result on statistics across entire sentences, which is widely used in conventional methods, STOI is based on shorter time segments [20].

The STOI algorithm compares a degraded signal with a clean original as reference. The output of the algorithm is in range $[0, 1]$ describing how much the degraded

signal corresponds to the original. Comparing the reference signal to itself will result in an output value of 1. In our case the input, before and after beamforming, is compared to the reference. An increased STOI-value, $\Delta_{STOI} > 0$, is viewed as an improvement after beamforming.

4.3.2 PESQ

Perceptual Evaluation of Speech Quality (PESQ) is an industry standard, consisting of methods for voice quality assessments. The method is widely used in the telecommunications industry and is put forward by The International Telecommunication Union (ITU) [21]. ITU also provides a reference implementation that is being used in this thesis [22]. The PESQ algorithm inputs two signals, one reference and one degraded. As an output it produces a value in the range $[-0.5, 4.5]$ describing how much the signal has degraded. A lower value means more degradation. The PESQ values are compared before and after beamforming. An increased value, $\Delta_{PESQ} > 0$, is viewed as an improvement.

Chapter 5

Results

This chapter contains information regarding the test results from the performed test sequences and describes the performance of the algorithms.

5.1 Direction of Arrival

The implemented method, as described in section 3.1, is equipped with a visual representation pointing out the DOA in real-time. Figure 5.1 shows the visual representation in different instances of time.

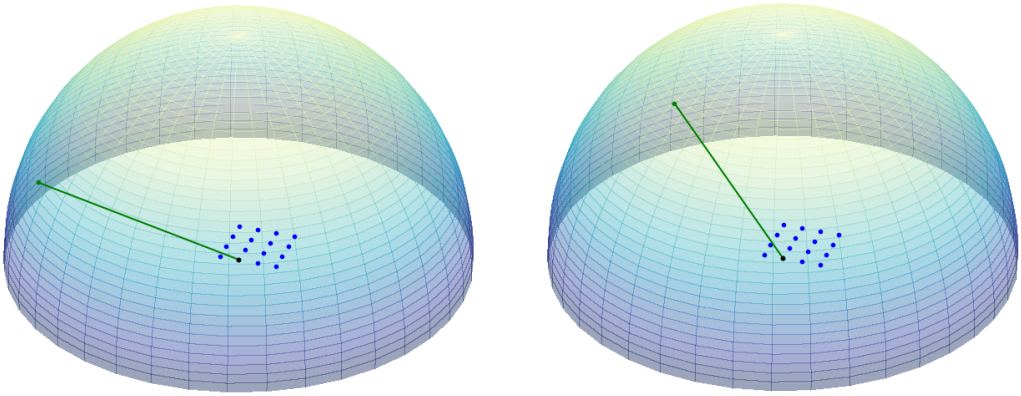


FIGURE 5.1: Visual representation of the DOA algorithm. Blue dots representing the microphone array with reference microphone in black. The green line represents the direction of arrival.

Due to prioritization of the BF algorithms no objective measurements were made. However, the algorithm was subjectively evaluated by introducing a SOI in a quiet office space. If a predominant SOI was present it was able to follow it with little or no fluctuations. But if another source was introduced the amount of fluctuations increased and it was more complicated to retrieve an unambiguous DOA.

Initially, the idea was to use the DOA from our algorithm and use it as input for our beamformer and in such way create a very basic BSS. However, due to instability, the algorithm was not suitable to couple with the beamformer.

5.2 Beamforming

The beamforming algorithms were applied to a total of 544 test permutations. Due to the large quantity of tables, only a handful are presented in the result section. In addition to beamforming, low-pass filtering was also applied for comparison. For the interested reader, all permutations and its results are presented in Appendix B.

Due to the precision of MVDR only a small mismatch of angle had a large negative impact on the results. Therefore, in order to assess the potential of MVDR, we decided to manually adjust the recorded SOI to be certain that it arrived from the correct direction.

5.2.1 Spectrogram

In order to intuitively see a difference in performance spectrograms are used. Spectrograms are visual representations of the frequency spectra over time of a signal. In Figure 5.2 below, a clean SOI is shown. This shows the frequencies (y-axis) present in the SOI in the recording over time (x-axis). The recordings are each 12 seconds long and consists of four different speaking parts, which also can be distinguished in the spectrogram below.

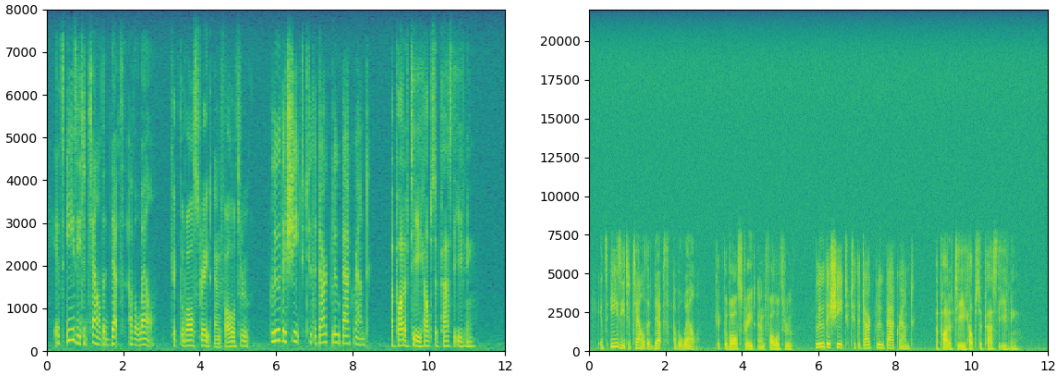


FIGURE 5.2: Spectrogram of clean SOI with $F_s = 16$ kHz and 44.1 kHz respectively.

Each segment represents one of the Harvard sentences, in the same order, from Table 4.2. Ultimately the goal is to recover the SOI and reproduce the spectrogram of the clean signal with use of beamforming.

5.2.2 Delay and Sum

Delay and sum is based on delaying the signal according to the DOA and then summing all the channels. Depending on the phase of the signal it will experience positive or negative interference. However, some frequencies have a wavelength corresponding to the TDOA, or the extra distance it has to travel from one to another

microphone, and will therefore experience positive interference and be gained instead. The effect in total is that the SOI will be evenly gained and noise coming from other directions will be attenuated.

Consider a SOI coming from an azimuth angle of 90 degrees and a elevation angle of 30 degrees. This puts the noise signals coming a from azimuth angle of 90 degrees and a elevation angle of 0, 15, 45 and 60 degrees respectively in accordance with Figure 4.3. All signals are delayed to have the SOI centered and then summed. This is referred to as putting the beam in direction $(\theta, \phi) = (90°, 30°)$. However, if the beam is put in the direction $(\theta, \phi) = (90°, x)$ where $x \in [0°, 90°]$, we will experience attenuation of the SOI as show in Figure 5.3 when $\phi \neq 30°$.

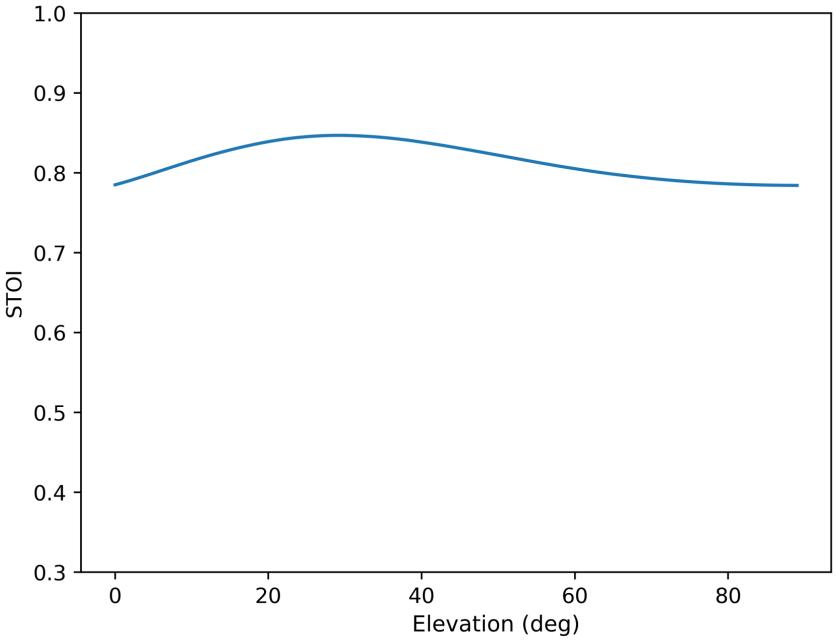


FIGURE 5.3: STOI results with fixed azimuth angle (90°).

The overall best test result with DS was obtained in the anechoic chamber, with a sample rate of 44.1 kHz, white noise as noise and fractional delay as shown in Table 5.1. The mean improvement of STOI (Δ_{STOI}) was 0.032 and of PESQ (Δ_{PESQ}) was 0.22. This setup also includes the best individual performance of STOI with an increase of 0.062. Table 5.1 also shows the performance of the rest of the fifteen angles under these conditions.

TABLE 5.1: Anechoic chamber, white noise as noise, with fractional delay, sample rate 44.1 kHz.

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.621	0.626	0.005	2.401	2.415	0.014
0	30	0.798	0.846	0.048	2.361	2.369	0.008
0	60	0.72	0.767	0.047	2.435	2.347	-0.088
0	90	0.697	0.711	0.015	0.632	2.314	**1.682**
30	0	0.852	0.853	0.002	2.499	2.307	-0.192
30	30	0.802	0.845	0.043	2.158	2.551	0.393
30	60	0.563	0.604	0.041	2.131	2.259	0.128
30	90	0.729	0.767	0.037	2.179	0.874	**-1.305**
60	0	0.852	0.851	**-0.001**	2.289	2.294	0.005
60	30	0.795	0.841	0.046	2.383	2.4	0.017
60	60	0.724	0.777	0.053	2.203	2.704	0.501
60	90	0.759	0.792	0.033	0.929	2.338	1.409
90	0	0.815	0.817	0.001	2.35	2.387	0.037
90	30	0.785	0.847	**0.062**	2.426	2.292	-0.134
90	60	0.714	0.773	0.059	1.762	2.302	0.54
90	90	0.783	0.805	0.022	1.825	2.326	0.501

To illustrate the improvement after BF a spectrogram is plotted of the considered signal from $(\theta, \phi) = (90^\circ, 30^\circ)$ as shown in Figure 5.4. The spectrogram to the left is the degraded signal before any processing, only a hint of the SOI can bee seen beneath 2.5 kHz. The spectrogram to the right shows the signal after processing. Now the SOI is more visible up to 7.5 kHz. The effect of full positive interference for frequencies with a wavelength corresponding to TDOA can be seen around 16 kHz where the noise is not attenuated as much as the rest.

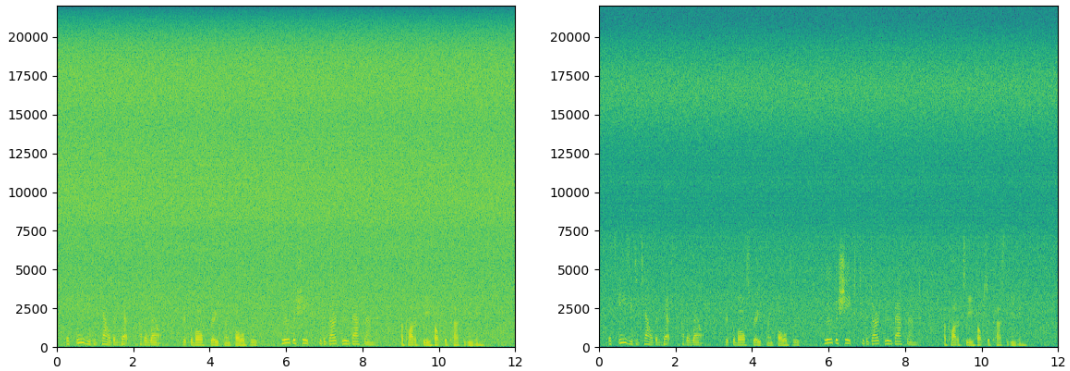


FIGURE 5.4: Spectrogram, before and after beamforming with DS respectively.

In figure 5.5 the STOI improvement Δ_{STOI} is plotted for delay-and-sum with, and without fractional delay together with the low-pass filter. Overall, it can be seen that the low-pass performs worse than both DS implementations. It can also be seen that the DS, with and without fractional delay, performs almost equally. This is expected,

as a fractional part is less than the sampling period and would only account for an extremely small change in the angle.

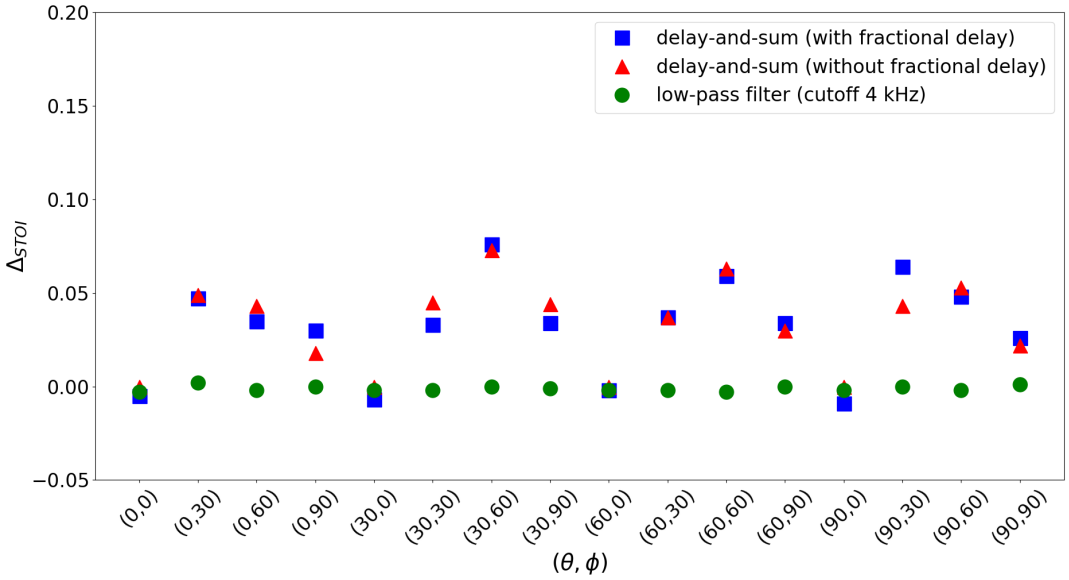


FIGURE 5.5: Anechoic chamber, white noise as noise, sample rate 16 kHz.

5.2.3 MVDR

Since MVDR is so meticulous, uncertainty in angling the microphone array was crucial to our test recordings. Even small errors in the DOA estimation have negative impact on the results. Due to the time limit on this thesis RMVB, which accounts for uncertainty in the angle, was not implemented. However, to demonstrate how powerful MVDR can be, the SOI was simulated to correspond to an angle without any uncertainty.

Given a SOI having a DOA from azimuth 90 and elevation 30 degrees, Figure 5.6 shows the result of placing the beam incorrectly. With a fixed azimuth of 90 degrees the beam is swept over all possible elevation angles, $\phi \in [0°, 90°]$. As can be seen the SOI will be suppressed when the beam is placed in the wrong direction. Recall from equation 2.27 that the output power is minimized without affecting the gain in a given direction. If the SOI is coming from another direction than the beam is pointing, it will subsequently be suppressed as if it were noise as can be seen in Figure 5.6.

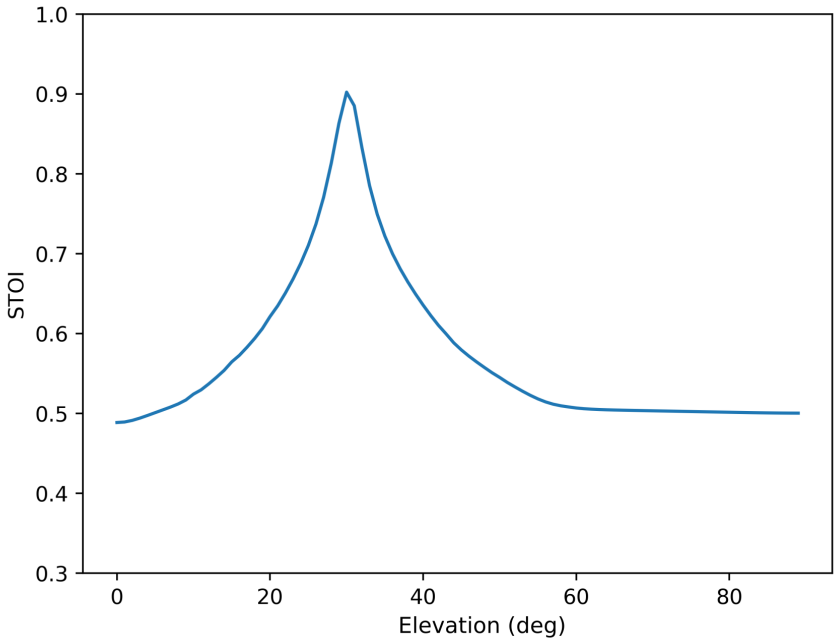


FIGURE 5.6: STOI results with fixed azimuth angle (90°).

The overall best test result was obtained in the anechoic chamber, with a sample rate of 16 kHz, white noise as noise and a 128-subband MVDR-beamformer as shown in Table 5.2. The mean improvement of STOI (Δ_{STOI}) was 0.074 and of PESQ (Δ_{PESQ}) was 0.54. This setup also includes the best individual performance of STOI with an increase of 0.153. Performance of the rest of the fifteen angles under these conditions can also be found in Table 5.2.

TABLE 5.2: Anechoic chamber, white noise as noise, 128 subbands, sample rate 16 kHz.

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.850	0.923	0.073	1.257	0.82	-0.437
0	30	0.753	0.905	0.153	1.39	1.205	-0.185
0	60	0.656	0.726	0.071	2.402	2.47	0.068
0	90	0.598	0.694	0.097	2.471	2.45	-0.021
30	0	0.849	0.927	0.078	1.399	2.459	1.06
30	30	0.767	0.833	0.066	1.731	2.447	0.716
30	60	0.647	0.665	0.018	1.195	2.47	1.275
30	90	0.603	0.602	**-0.001**	0.462	2.479	**2.017**
60	0	0.840	0.922	0.082	2.426	2.454	0.028
60	30	0.755	0.814	0.059	1.987	1.515	**-0.472**
60	60	0.674	0.718	0.044	1.498	2.997	1.499
60	90	0.613	0.652	0.039	1.69	2.465	0.775
90	0	0.838	0.921	0.083	1.378	1.323	-0.055
90	30	0.749	0.902	**0.153**	1.474	2.464	0.99
90	60	0.697	0.758	0.061	1.135	2.473	1.338
90	90	0.646	0.759	0.113	2.398	2.46	0.062

To illustrate the improvement after BF a spectrogram is plotted of the considered signal from $(\theta, \phi) = (90°, 30°)$ as shown in Figure 5.7. The spectrogram to the left is the degraded signal before any processing, only a hint of the SOI can bee seen beneath 1.5 kHz. The spectrogram to the right shows the signal after processing. Now the SOI is more visible up to 7.5 kHz. Since MVDR is performed in subbands the frequencies are evenly attenuated.

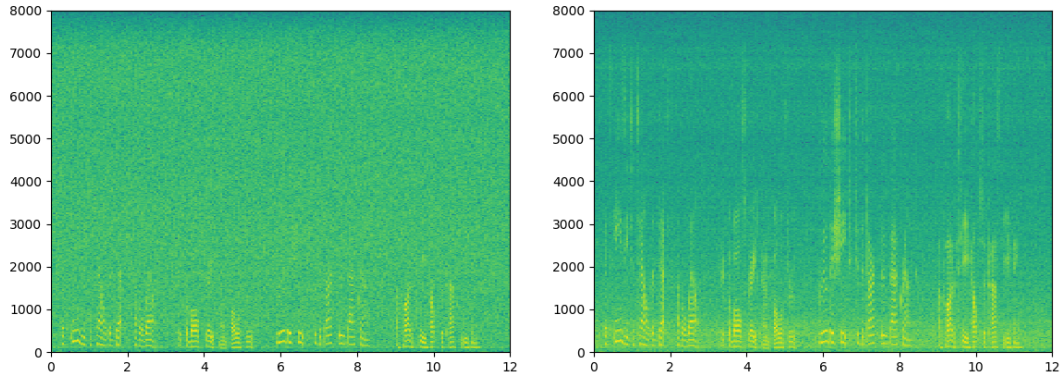


FIGURE 5.7: Spectrogram, before and after beamforming with MVDR respectively.

All spectrograms up till this point are from when white noise was used as noise source. For a more intuitive picture of what MVDR is capable of, a spectrogram is added with tones as noise source, as seen in Figure 5.8. The setup illustrated is from the anechoic chamber with a sample frequency of 16 kHz from an angle of $(\theta, \phi) = (60°, 90°)$ and 128 subbands used for processing with MVDR. Recall from section 4.2 that the tones used as noise are $0.5, 1, 2, 3$ kHz respectively. The lines corresponding to these frequencies are showing in the lower part of the figure. The rest of the lines are a result of overtones present in each of the sinusoidal signals. As shown to the right, MVDR manages to suppress the tones and retrieve more SOI information from the frequencies above 2 kHz.

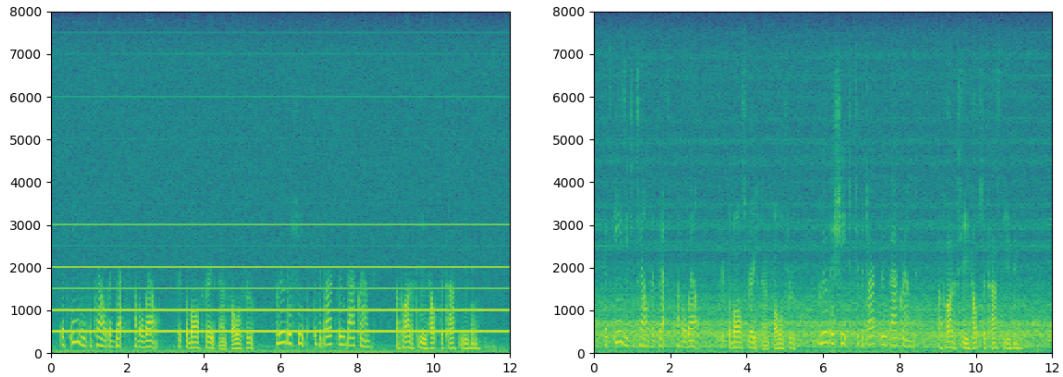


FIGURE 5.8: Spectrogram, before and after beamforming with MVDR respectively.

To summarize the results above, the mean STOI improvements of the different methods are presented in Table 5.3. The setup used for comparison is the anechoic chamber with white noise as noise, a sample rate of 16 kHz (with 128 subbands for MVDR).

Method	$\overline{\Delta}_{STOI}$
MVDR	0.074
DS (with fractional delay)	0.031
DS (without fractional delay)	0.033
Low-pass filter, cutoff 4 kHz	-0.001

TABLE 5.3: $\overline{\Delta}_{STOI}$ of performed tests obtained from Table B.15, B.23, B.25 and B.33 respectively.

Chapter 6

Conclusion and Future Work

In this chapter we discuss limitations, possible improvements and present our conclusions. The chapter also contains some suggestions for future work.

6.1 Conclusion

The objective of this thesis was to discuss different DOA and BF techniques and subsequently evaluate these algorithms after implementing them. Based on our findings the MVDR beamformer shows great potential for use in beamforming whereas the delay-and-sum beamformer shows less capability. Although MVDR shows potential it is still not suitable for some use cases in real applications due to the sensitivity with the direction of arrival of the incident waves.

For test evaluation some of the results are contradictory between STOI and PESQ. In some cases there is an increase in STOI while PESQ is decreased and vice versa. However, a subjective listening evaluation indicates that there is an improvement in speech perception, also for the results suggesting degradation. As a consequence it is recommended to refine the criteria of units of measurement but still make use of subjective listening while evaluating microphone array processing methods in the future.

6.2 Future Work

Based on the work done throughout this thesis a couple of suggestions on future work are given below. Discussions already occurred during the thesis whether to add the areas in our work or not, but unfortunately the time frame did not allow us to pursue this.

Test Data Set

In total 256 test sequences were recorded during this thesis. The data set is in its original state and not modified. This allows for support in testing and evaluating different or improved algorithms in the future.

Microphone Array Setups

The thesis was limited to one commercially available microphone array setup only. Planar, 16 microphones and equally spaced with 4.2 cm apart. As stated in section 2.3.2 this gave us the best suppression of side lobes as well as a narrow main lobe in comparison to 4 or 9 microphones respectively. How well would other array geometry, e.g circular or triangular, affect the results of noise suppression? Would the result be better if even more microphones were used or would it just make the task infeasible due to computational complexity?

RMVB

Due to the precision of the MVDR beamformer it is currently not suitable for use in real world environments since the uncertainty of the direction of arrival is too large in most use cases. It is desirable to have better robustness so that the SOI does not get suppressed. A solution for this is the RMVB which is an extension of MVDR which accounts for uncertainty in the DOA.

Real-time Implementation

We were only able to evaluate our DOA algorithm in real-time. For beamforming the amount of computations were too many and the program would shut down quickly. A real-time application of beamforming would allow a wide range of use cases. But for this to work, investigation of how to decrease the computational complexity and how much computational power is needed is a necessity.

Blind Signal Separation

A wide area that has not been discussed enough is blind signal separation. It is a method that can be used to separate a set of mixed signals with no or little information about the sources or about how they were mixed [23]. Ideally, and as we initially planned, one could investigate, implement and compare this method with a combination of DOA and BF.

Appendix A

UMA-16 Microphone Array

UMA-16 Microphone Array

Features
- 16channel Microphone Array
- Uniform Rectangular Array (URA)
- USB, PDM & Customizable IO card

Hardware
- ADI ADSP21489 @ 400MHz
- XMOS XCore200 @ 500MHz
- 16 SPH1668LM4H Knowles MEMS
- Asynchronous USB audio
- Mic array PCB schematics are provided as a reference design

Software Control
- ASIO drivers for Win 7/8/10
- Driverless UAC2 for Mac OSx/linux
- Compatible with Matlab toolbox
- Firmware upgradeable

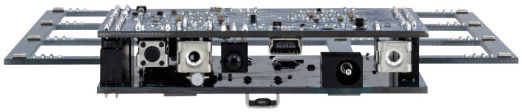
Power
- Single external 12VDC supply

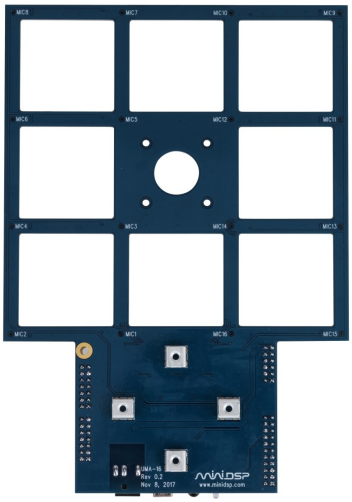
Applications
- Acoustic camera
- Research & Development
- Microphone array

The **UMA-16** is a sixteen channels microphone array with plug&play USB audio connectivity. With its onboard SHARC+XMOS controller board, the **UMA-16** is the perfect fit for the development of beamforming algorithms or your DIY acoustic camera. Its system architecture consists of two core elements:

- The microphone array PCB has 16 x SPH1668LM4H MEMS Knowles laid out in a Uniform Rectangular Array (URA). A center hole fits an optional USB camera for applications such as an acoustic camera. Being a simple 2layer PCN, one can easily customize his own array layout by following our schematics.

- At the helm of the **UMA-16** operation is the nanoSharc kit. A 400MHz SHARC ADSP21489 + 500MHz multicore CPU providing substantial processing power for high SNR PDM to PCM conversion and multichannel low latency USB audio.

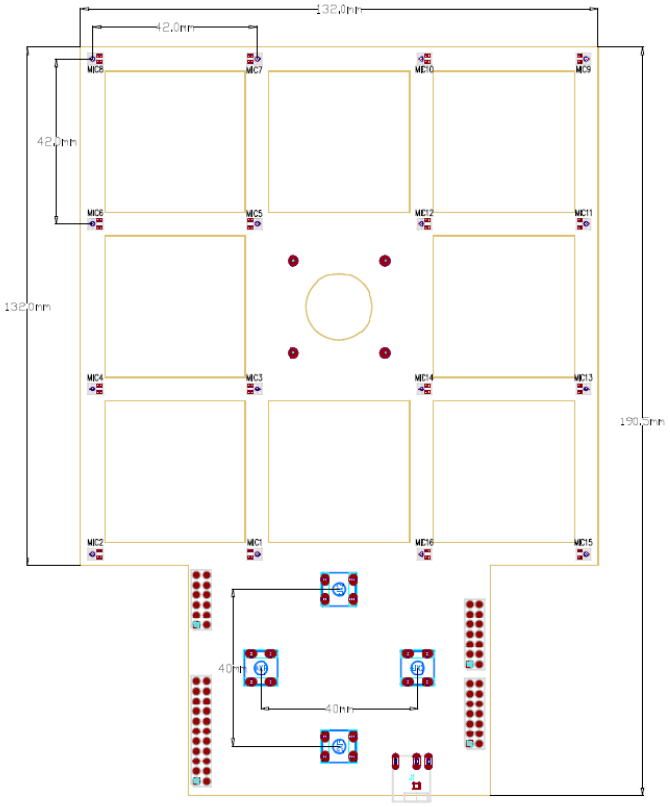




TECHNICAL SPECIFICATIONS

Item	Description
Digital Signal Processor	32-bit Floating point Analog Devices SHARC ADSP21489 / 400 MHz - Configuration locked
USB audio input	XMOS Xcore200 asynchronous USB audio up to 192 kHz, USB Audio Class 2 compliant . ASIO drivers for Windows . Driverless for Mac OS X
PDM inputs	Up to 16 x MEMS microphone connections (8 x stereo PDM data lines)
MEMS microphone	16 x SPH1668LM4H - Acoustic Overload @ 120dBSPL / High SNR of 65dB / RF shielded
ADC/DAC Sample rate & Resolution	Resolution: 24 bit Sample rate: 14.7k/11.025k/12k/16k/22.05k/44.1k/48k
USB port	USB port type Mini-B for audio streaming and firmware upgrade
Power supply	12 VDC single supply / Header input / 2.5W
Dimensions (H x W x D) mm	132 x 195 x 25 mm
Mounting	4 x M3 holders for front panel mounting / CAD drawings available on demand

MECHANICAL DRAWING



J3 Header - 11x2			
Usage	**Pin Number**		**Usage**
Not in use	1	2	Not in use
Not in use	3	4	Not in use
Not in use	5	6	Do Not Connect
GND	7	8	PDM[0]
PDM[1]	9	10	PDM[2]
PDM[3]	11	12	PDM[4]
PDM[5]	13	14	PDM[6]
PDM[7]	15	16	PDM CLK1
PDM CLK2	17	18	Not in use
GND	19	20	GND
3V3	21	22	3V3

J2 Header - 6x2			
Usage	**Pin Number**		**Usage**
I2S LRCLK	1	2	I2S BCLK
GND	3	4	I2S MCLK
I2S Out0	5	6	Not in use
I2C SCLK	7	8	I2C SDA
GND	9	10	GND
12V+ IN	11	12	12V+ IN

Appendix B

Test Results

This appendix contains all test results for every test sequence run. For each setup the best and worst improvement of both STOI and PESQ are highlighted with green and red color respectively.

B.1 MVDR

TABLE B.1: Anechoic chamber, tones as noise, 128 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.889	0.946	0.057	1.271	1.643	0.372
0	30	0.856	0.941	0.086	1.492	1.882	0.39
0	60	0.786	0.834	0.047	1.149	2.547	**1.398**
0	90	0.716	0.806	0.090	1.543	0.656	-0.887
30	0	0.879	0.945	0.066	2.405	1.503	-0.902
30	30	0.875	0.831	**-0.043**	1.668	2.466	0.798
30	60	0.737	0.786	0.049	1.902	2.46	0.558
30	90	0.664	0.752	0.088	2.383	2.465	0.082
60	0	0.875	0.944	0.069	2.592	2.595	0.003
60	30	0.853	0.821	-0.033	1.498	2.434	0.936
60	60	0.773	0.832	0.059	0.392	1.12	0.728
60	90	0.667	0.806	**0.139**	4.23	2.462	**-1.768**
90	0	0.891	0.942	0.051	2.419	1.003	-1.416
90	30	0.851	0.933	0.082	1.474	2.309	0.835
90	60	0.784	0.817	0.033	2.428	1.385	-1.043
90	90	0.758	0.862	0.104	0.662	1.68	1.018

TABLE B.2: Anechoic chamber, tones as noise, 256 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.889	0.932	0.043	1.271	2.457	1.186
0	30	0.856	0.928	0.072	1.492	1.103	-0.389
0	60	0.786	0.820	0.034	1.149	1.351	0.202
0	90	0.716	0.803	0.087	1.543	0.816	-0.727
30	0	0.879	0.933	0.054	2.405	2.464	0.059
30	30	0.875	0.827	**-0.048**	1.668	2.456	0.788
30	60	0.737	0.786	0.049	1.902	2.452	0.55
30	90	0.664	0.752	0.088	2.383	1.607	-0.776
60	0	0.875	0.933	0.058	2.592	2.463	-0.129
60	30	0.853	0.811	-0.042	1.987	0.785	-1.202
60	60	0.773	0.824	0.051	0.392	2.468	**2.076**
60	90	0.667	0.799	**0.132**	4.23	2.453	**-1.777**
90	0	0.891	0.929	0.038	2.419	2.456	0.037
90	30	0.851	0.920	0.069	1.474	1.893	0.419
90	60	0.784	0.807	0.023	2.428	1.885	-0.543
90	90	0.758	0.855	0.096	0.662	2.467	1.805

TABLE B.3: Anechoic chamber, tones as noise, 512 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.889	0.912	0.023	1.271	0.445	-0.826
0	30	0.856	0.913	0.057	1.492	1.021	-0.471
0	60	0.786	0.806	0.020	1.149	2.471	**1.322**
0	90	0.716	0.788	0.072	1.543	2.467	0.924
30	0	0.879	0.910	0.031	2.405	1.943	-0.462
30	30	0.875	0.814	**-0.060**	1.668	2.125	0.457
30	60	0.737	0.776	0.039	1.902	2.469	0.567
30	90	0.664	0.745	0.081	2.383	2.476	0.093
60	0	0.875	0.914	0.040	2.592	1.486	-1.106
60	30	0.853	0.799	-0.054	1.498	1.89	0.392
60	60	0.773	0.814	0.041	0.392	1.206	0.814
60	90	0.667	0.791	**0.124**	4.23	2.468	**-1.762**
90	0	0.891	0.909	0.018	2.419	2.461	0.042
90	30	0.851	0.903	0.052	1.474	2.566	1.092
90	60	0.784	0.797	0.012	2.428	1.045	-1.383
90	90	0.758	0.839	0.081	0.662	1.994	**1.332**

TABLE B.4: Anechoic chamber, tones as noise, 128 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.882	0.986	0.103	2.2	2.329	0.129
0	30	0.858	0.919	0.061	0.179	2.33	**2.151**
0	60	0.804	0.863	0.059	2.35	2.329	-0.021
0	90	0.729	0.809	0.080	1.4	2.339	0.939
30	0	0.878	0.985	0.107	2.291	2.334	0.043
30	30	0.877	0.892	**0.015**	2.212	0.741	**-1.471**
30	60	0.768	0.843	0.074	1.577	2.321	0.744
30	90	0.691	0.821	**0.130**	2.313	2.352	0.039
60	0	0.902	0.985	0.083	2.326	2.335	0.009
60	30	0.863	0.880	0.017	0.25	1.979	1.729
60	60	0.775	0.868	0.093	2.02	2.34	0.32
60	90	0.707	0.835	0.129	1.108	2.346	1.238
90	0	0.876	0.965	0.090	2.281	1.816	-0.465
90	30	0.858	0.880	0.022	1.166	2.329	1.163
90	60	0.795	0.872	0.077	2.214	2.322	0.108
90	90	0.774	0.873	0.099	2.168	2.348	0.18

TABLE B.5: Anechoic chamber, tones as noise, 256 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.882	0.980	0.098	2.2	1.435	-0.765
0	30	0.858	0.895	0.037	0.179	2.322	**2.143**
0	60	0.804	0.848	0.044	2.35	2.331	-0.019
0	90	0.729	0.807	0.079	1.4	2.339	0.939
30	0	0.878	0.982	0.104	2.291	2.325	0.034
30	30	0.877	0.872	-0.005	2.212	2.324	0.112
30	60	0.768	0.819	0.051	1.577	0.403	-1.174
30	90	0.691	0.808	**0.117**	2.313	2.34	0.027
60	0	0.902	0.980	0.078	2.326	2.326	0
60	30	0.863	0.857	-0.006	0.25	2.324	2.074
60	60	0.775	0.859	0.084	2.02	2.328	0.308
60	90	0.707	0.818	0.111	1.108	2.347	1.239
90	0	0.876	0.958	0.082	2.281	0.184	**-2.097**
90	30	0.858	0.843	**-0.016**	1.166	2.317	1.151
90	60	0.795	0.839	0.044	2.214	2.317	0.103
90	90	0.774	0.834	0.060	2.168	2.351	0.183

TABLE B.6: Anechoic chamber, tones as noise, 512 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.882	0.974	0.092	2.2	0.827	-1.373
0	30	0.858	0.889	0.031	0.179	1.897	1.718
0	60	0.804	0.846	0.042	2.35	0.924	-1.426
0	90	0.729	0.815	0.086	1.4	2.34	0.94
30	0	0.878	0.974	0.096	2.291	1.768	-0.523
30	30	0.877	0.867	-0.010	2.212	2.317	0.105
30	60	0.768	0.821	0.052	1.577	0.249	-1.328
30	90	0.691	0.820	**0.129**	2.313	2.341	0.028
60	0	0.902	0.973	0.071	2.326	2.331	0.005
60	30	0.863	0.852	-0.010	0.25	2.333	**2.083**
60	60	0.775	0.864	0.089	2.02	2.367	0.347
60	90	0.707	0.825	0.118	1.108	2.35	1.242
90	0	0.876	0.948	0.073	2.281	2.303	0.022
90	30	0.858	0.839	**-0.019**	1.166	2.307	1.141
90	60	0.795	0.838	0.043	2.214	2.323	0.109
90	90	0.774	0.840	0.066	2.168	0.062	**-2.106**

TABLE B.7: Anechoic chamber, white noise as noise, 128 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.850	0.923	0.073	1.257	0.82	-0.437
0	30	0.753	0.905	0.153	1.39	1.205	-0.185
0	60	0.656	0.726	0.071	2.402	2.47	0.068
0	90	0.598	0.694	0.097	2.471	2.45	-0.021
30	0	0.849	0.927	0.078	1.399	2.459	1.06
30	30	0.767	0.833	0.066	1.731	2.447	0.716
30	60	0.647	0.665	0.018	1.195	2.47	1.275
30	90	0.603	0.602	**-0.001**	0.462	2.479	**2.017**
60	0	0.840	0.922	0.082	2.426	2.454	0.028
60	30	0.755	0.814	0.059	1.987	1.515	**-0.472**
60	60	0.674	0.718	0.044	1.498	2.997	1.499
60	90	0.613	0.652	0.039	1.69	2.465	0.775
90	0	0.838	0.921	0.083	1.378	1.323	-0.055
90	30	0.749	0.902	**0.153**	1.474	2.464	0.99
90	60	0.697	0.758	0.061	1.135	2.473	1.338
90	90	0.646	0.759	0.113	2.398	2.46	0.062

TABLE B.8: Anechoic chamber, white noise as noise, 256 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.850	0.916	0.065	1.257	2.466	1.209
0	30	0.753	0.903	**0.151**	1.39	1.069	-0.321
0	60	0.656	0.730	0.075	2.402	1.835	-0.567
0	90	0.598	0.704	0.106	2.471	2.703	0.232
30	0	0.849	0.920	0.070	1.399	1.009	-0.39
30	30	0.767	0.828	0.062	1.731	2.465	0.734
30	60	0.647	0.670	0.024	1.195	1.137	-0.058
30	90	0.603	0.608	**0.005**	0.462	2.483	**2.021**
60	0	0.840	0.913	0.073	2.426	2.454	0.028
60	30	0.755	0.809	0.054			
60	60	0.674	0.729	0.055	1.498	2.459	0.961
60	90	0.613	0.670	0.056	1.69	2.453	0.763
90	0	0.838	0.913	0.075	1.378	1.098	-0.28
90	30	0.749	0.898	0.149	1.474	0.479	-0.995
90	60	0.697	0.767	0.070	1.135	1.89	0.755
90	90	0.646	0.772	0.125	2.398	0.21	**-2.188**

TABLE B.9: Anechoic chamber, white noise as noise, 512 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.850	0.901	0.051	1.257	1.614	0.357
0	30	0.753	0.894	**0.141**	1.39	1.206	-0.184
0	60	0.656	0.735	0.080	2.402	2.484	0.082
0	90	0.598	0.708	0.110	2.471	1.463	**-1.008**
30	0	0.849	0.905	0.056	1.399	2.005	0.606
30	30	0.767	0.821	0.054	1.731	2.473	0.742
30	60	0.647	0.676	0.030	1.195	2.461	1.266
30	90	0.603	0.619	**0.016**	0.462	2.473	**2.011**
60	0	0.840	0.900	0.060	2.426	2.459	0.033
60	30	0.755	0.801	0.045	1.987	1.62	-0.367
60	60	0.674	0.733	0.059	1.498	1.062	-0.436
60	90	0.613	0.678	0.065	1.69	1.07	-0.62
90	0	0.838	0.898	0.060	1.378	2.632	1.254
90	30	0.749	0.888	0.140	1.474	2.457	0.983
90	60	0.697	0.766	0.069	1.135	2.319	1.184
90	90	0.646	0.775	0.129	2.398	2.471	0.073

TABLE B.10: Anechoic chamber, white noise as noise, 128 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.856	0.956	0.100	2.397	2.315	-0.082
0	30	0.758	0.870	**0.112**	1.598	2.331	0.733
0	60	0.657	0.718	0.061	2.425	2.342	-0.083
0	90	0.602	0.651	0.050	0.232	2.34	**2.108**
30	0	0.851	0.960	0.108	2.317	1.139	-1.178
30	30	0.768	0.859	0.091	2.392	2.326	-0.066
30	60	0.681	0.697	0.015	1.826	0.046	-1.78
30	90	0.644	0.637	**-0.006**	2.284	2.328	0.044
60	0	0.852	0.953	0.101	2.344	0.181	**-2.163**
60	30	0.764	0.854	0.090	2.119	2.339	0.22
60	60	0.660	0.736	0.076	2.362	2.323	-0.039
60	90	0.657	0.665	0.008	2.212	2.341	0.129
90	0	0.822	0.925	0.103	2.49	2.362	-0.128
90	30	0.753	0.836	0.083	2.426	2.159	-0.267
90	60	0.690	0.769	0.080	2.332	2.327	-0.005
90	90	0.681	0.740	0.059	2.102	2.347	0.245

TABLE B.11: Anechoic chamber, white noise as noise, 256 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.856	0.956	0.100	2.397	1.965	-0.432
0	30	0.758	0.872	**0.114**	1.598	1.755	0.157
0	60	0.657	0.725	0.067	2.425	1.73	-0.695
0	90	0.602	0.665	0.063	0.232	1.738	**1.506**
30	0	0.851	0.961	0.109	2.317	2.557	0.24
30	30	0.768	0.853	0.085	2.392	2.321	-0.071
30	60	0.681	0.685	0.004	1.826	1.705	-0.121
30	90	0.644	0.626	**-0.017**	2.284	2.324	0.04
60	0	0.852	0.952	0.100	2.344	2.325	-0.019
60	30	0.764	0.839	0.075	2.119	2.338	0.219
60	60	0.660	0.745	0.086	2.362	0.598	**-1.764**
60	90	0.657	0.647	-0.010	2.212	2.348	0.136
90	0	0.822	0.926	0.103	2.49	2.324	-0.166
90	30	0.753	0.816	0.063	2.426	2.332	-0.094
90	60	0.690	0.760	0.071	2.332	2.327	-0.005
90	90	0.681	0.720	0.039	2.102	0.624	-1.478

TABLE B.12: Anechoic chamber, white noise as noise, 512 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.856	0.953	0.098	2.397	1.447	-0.95
0	30	0.758	0.877	**0.119**	1.598	1.219	-0.379
0	60	0.657	0.753	0.095	2.425	1.526	-0.899
0	90	0.602	0.693	0.091	0.232	0.858	**0.626**
30	0	0.851	0.960	0.109	2.317	0.832	-1.485
30	30	0.768	0.855	0.087	2.392	2.356	-0.036
30	60	0.681	0.708	0.027	1.826	2.234	0.408
30	90	0.644	0.652	**0.008**	2.284	1.027	-1.257
60	0	0.852	0.952	0.101	2.344	2.334	-0.01
60	30	0.764	0.839	0.075	2.119	0.4	**-1.719**
60	60	0.660	0.761	0.101	2.362	2.328	-0.034
60	90	0.657	0.679	0.022	2.212	1.425	-0.787
90	0	0.822	0.924	0.102	2.49	2.316	-0.174
90	30	0.753	0.822	0.069	2.426	2.322	-0.104
90	60	0.690	0.778	0.088	2.332	1.739	-0.593
90	90	0.681	0.748	0.066	2.102	0.407	-1.695

TABLE B.13: Studio, white noise as noise, 128 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.830	0.870	0.041	2.091	1.257	-0.834
0	30	0.738	0.842	**0.104**	2.544	1.383	**-1.161**
0	60	0.676	0.687	0.011	2.582	2.473	-0.109
0	90	0.711	0.804	0.093	1.29	2.48	**1.19**
30	0	0.847	0.885	0.037	1.603	2.48	0.877
30	30	0.764	0.731	-0.033	1.304	1.904	0.6
30	60	0.703	0.678	-0.025	1.346	0.68	-0.666
30	90	0.653	0.635	-0.017	1.963	2.478	0.515
60	0	0.827	0.863	0.036	2.85	2.543	-0.307
60	30	0.738	0.695	**-0.043**	1.492	1.917	0.425
60	60	0.701	0.666	-0.034	1.444	2.478	1.034
60	90	0.664	0.672	0.008	1.444	2.476	1.032
90	0	0.818	0.859	0.041	2.193	1.508	-0.685
90	30	0.746	0.819	0.073	2.053	1.211	-0.842
90	60	0.707	0.675	-0.031	2.419	2.481	0.062
90	90	0.656	0.669	0.013	2.583	2.482	-0.101

TABLE B.14: Studio, white noise as noise, 256 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.830	0.864	0.034	2.091	1.733	-0.358
0	30	0.738	0.850	0.112	2.544	2.492	-0.052
0	60	0.676	0.719	0.042	2.582	0.434	**-2.148**
0	90	0.711	0.835	**0.124**	1.29	2.256	0.966
30	0	0.847	0.876	0.029	1.603	1.542	-0.061
30	30	0.764	0.729	-0.035	1.304	1.898	0.594
30	60	0.703	0.700	-0.003	1.346	2.472	**1.126**
30	90	0.653	0.655	0.002	1.963	2.473	0.51
60	0	0.827	0.856	0.029	2.85	2.47	-0.38
60	30	0.738	0.692	**-0.046**	1.492	2.476	0.984
60	60	0.701	0.689	-0.012	1.444	1.8	0.356
60	90	0.664	0.692	0.029	1.444	0.428	-1.016
90	0	0.818	0.850	0.031	2.193	1.842	-0.351
90	30	0.746	0.823	0.077	2.053	2.462	0.409
90	60	0.707	0.693	-0.014	2.419	1.493	-0.926
90	90	0.656	0.699	0.043	2.583	2.493	-0.09

TABLE B.15: Studio, white noise as noise, 512 subbands, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.830	0.847	0.017			
0	30	0.738	0.836	0.098	2.091	0.981	-1.11
0	60	0.676	0.717	0.041	2.544	2.486	-0.058
0	90	0.711	0.838	**0.127**	2.582	0.423	**-2.159**
30	0	0.847	0.856	0.008	1.29	2.479	**1.189**
30	30	0.764	0.714	-0.050	1.603	0.437	-1.166
30	60	0.703	0.701	-0.002	1.304	2.481	1.177
30	90	0.653	0.665	0.012	1.346	2.483	1.137
60	0	0.827	0.836	0.009	1.963	1.031	-0.932
60	30	0.738	0.677	**-0.061**	2.85	2.482	-0.368
60	60	0.701	0.684	-0.017	1.492	2.477	0.985
60	90	0.664	0.695	0.031	1.444	1.762	0.318
90	0	0.818	0.831	0.012	1.444	2.47	1.026
90	30	0.746	0.804	0.058	2.053	1.176	-0.877
90	60	0.707	0.685	-0.021	2.419	2.493	0.074
90	90	0.656	0.709	0.053	2.583	2.485	-0.098

TABLE B.16: Studio, white noise as noise, 128 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.885	0.945	0.061	2.425	2.368	-0.057
0	30	0.750	0.758	0.008	2.52	2.375	-0.145
0	60	0.687	0.670	-0.016	2.394	1.156	-1.238
0	90	0.727	0.724	-0.003	2.333	1.218	-1.115
30	0	0.876	0.934	0.057	2.311	1.514	-0.797
30	30	0.769	0.750	-0.020	2.376	2.357	-0.019
30	60	0.724	0.673	-0.051	2.181	1.385	-0.796
30	90	0.633	0.602	-0.031	1.412	0.889	-0.523
60	0	0.865	0.928	**0.063**	2.719	0.157	**-2.562**
60	30	0.801	0.763	-0.037	1.405	1.855	0.45
60	60	0.715	0.693	-0.022	2.314	2.37	0.056
60	90	0.696	0.678	-0.018	2.11	0.632	-1.478
90	0	0.828	0.884	0.056	2.418	1.532	-0.886
90	30	0.782	0.732	-0.050	0.281	2.356	**2.075**
90	60	0.734	0.665	**-0.068**	1.748	1.472	-0.276
90	90	0.693	0.629	-0.064	2.414	2.37	-0.044

TABLE B.17: Studio, white noise as noise, 256 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.885	0.950	**0.065**	2.425	0.181	**-2.244**
0	30	0.750	0.758	0.007	2.52	2.365	-0.155
0	60	0.687	0.682	-0.005	2.394	2.357	-0.037
0	90	0.727	0.727	0.001	2.333	2.368	0.035
30	0	0.876	0.940	0.063	2.311	2.352	0.041
30	30	0.769	0.727	-0.042	2.376	2.358	-0.018
30	60	0.724	0.642	-0.081	2.181	2.364	0.183
30	90	0.633	0.579	-0.054	1.412	1.811	0.399
60	0	0.865	0.930	0.065	2.719	1.39	-1.329
60	30	0.801	0.722	-0.079	1.405	1.748	0.343
60	60	0.715	0.680	-0.035	2.314	0.58	-1.734
60	90	0.696	0.648	-0.048	2.11	1.062	-1.048
90	0	0.828	0.891	0.063	2.418	0.273	-2.145
90	30	0.782	0.712	-0.070	0.281	2.365	**2.084**
90	60	0.734	0.648	**-0.086**	1.748	2.353	0.605
90	90	0.693	0.615	-0.078	2.414	2.372	-0.042

TABLE B.18: Studio, white noise as noise, 512 subbands, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.885	0.955	**0.070**	2.425	1.091	-1.334
0	30	0.750	0.769	0.019	2.52	1.35	-1.17
0	60	0.687	0.723	0.037	2.394	2.361	-0.033
0	90	0.727	0.777	0.050	2.333	2.362	0.029
30	0	0.876	0.941	0.064	2.311	2.364	0.053
30	30	0.769	0.733	-0.036	2.376	2.351	-0.025
30	60	0.724	0.654	-0.070	2.181	2.364	0.183
30	90	0.633	0.609	-0.024	1.412	2.366	0.954
60	0	0.865	0.932	0.067	2.719	0.937	-1.782
60	30	0.801	0.717	**-0.083**	1.405	2.356	0.951
60	60	0.715	0.718	0.003	2.314	2.358	0.044
60	90	0.696	0.677	-0.019	2.11	2.355	0.245
90	0	0.828	0.898	**0.070**	2.418	0.264	**-2.154**
90	30	0.782	0.699	-0.083	0.281	2.358	**2.077**
90	60	0.734	0.677	-0.057	1.748	2.357	0.609
90	90	0.693	0.649	-0.044	2.414	2.389	-0.025

B.2 Delay and Sum

TABLE B.19: Studio, white noise as noise, with fractional delay, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.836	0.837	0.001	1.473	2.353	0.88
0	30	0.533	0.535	0.002	0.765	2.429	1.664
0	60	0.533	0.552	0.019	1.191	2.482	1.291
0	90	0.542	0.539	-0.002	2.466	1.431	-1.035
30	0	0.607	0.621	0.014	2.46	1.652	-0.808
30	30	0.561	0.553	**-0.008**	1.423	1.829	0.406
30	60	0.783	0.821	**0.038**	2.488	2.446	-0.042
30	90	0.517	0.518	0.002	2.537	1.255	**-1.282**
60	0	0.663	0.682	0.018	2.598	2.604	0.006
60	30	0.539	0.534	-0.005	0.829	2.424	1.595
60	60	0.562	0.58	0.018	2.408	2.451	0.043
60	90	0.514	0.519	0.005	2.459	2.458	-0.001
90	0	0.583	0.598	0.015	2.475	2.482	0.007
90	30	0.546	0.544	-0.003	1.174	1.418	0.244
90	60	0.513	0.53	0.017	0.98	2.655	**1.675**
90	90	0.511	0.511	0.001	2.454	2.483	0.029

TABLE B.20: Studio, white noise as noise, no fractional delay, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.836	0.836	0.0	1.473	1.474	0.001
0	30	0.533	0.55	0.017	0.765	2.44	**1.675**
0	60	0.533	0.537	0.004	1.191	1.154	-0.037
0	90	0.542	0.559	0.017	2.466	1.301	-1.165
30	0	0.607	0.607	-0.0	2.46	2.459	-0.001
30	30	0.561	0.57	0.01	1.423	2.575	1.152
30	60	0.783	0.822	**0.039**	2.488	2.46	-0.028
30	90	0.517	0.505	**-0.012**	2.537	0.732	**-1.805**
60	0	0.663	0.663	-0.0	2.598	2.599	0.001
60	30	0.539	0.549	0.009	0.829	1.939	1.11
60	60	0.562	0.565	0.004	2.408	2.483	0.075
60	90	0.514	0.536	0.022	2.459	2.484	0.025
90	0	0.583	0.583	0.0	2.475	2.629	0.154
90	30	0.546	0.561	0.014	1.174	0.534	-0.64
90	60	0.513	0.518	0.005	0.98	2.572	1.592
90	90	0.511	0.529	0.019	2.454	2.478	0.024

TABLE B.21: Studio, white noise as noise, with fractional delay, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.836	0.844	0.008	2.207	2.403	0.196
0	30	0.78	0.828	**0.048**	0.988	0.819	-0.169
0	60	0.504	0.528	0.024	2.509	2.266	-0.243
0	90	0.579	0.571	**-0.008**	2.426	2.376	-0.05
30	0	0.582	0.576	-0.006	2.347	2.375	0.028
30	30	0.573	0.591	0.018	1.521	2.18	**0.659**
30	60	0.793	0.836	0.043	1.887	2.341	0.454
30	90	0.734	0.733	-0.001	2.25	2.148	-0.102
60	0	0.597	0.591	-0.006	2.304	2.277	-0.027
60	30	0.6	0.624	0.024	2.101	2.652	0.551
60	60	0.486	0.506	0.02	0.32	0.063	-0.257
60	90	0.563	0.558	-0.005	2.389	1.999	-0.39
90	0	0.561	0.556	-0.006	2.495	2.475	-0.02
90	30	0.523	0.536	0.013	2.436	1.574	-0.862
90	60	0.493	0.51	0.017	2.31	1.395	**-0.915**
90	90	0.545	0.548	0.004	2.426	2.507	0.081

TABLE B.22: Studio, white noise as noise, no fractional delay, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.836	0.836	0.0	2.207	2.202	-0.005
0	30	0.78	0.828	**0.048**	0.988	2.383	**1.395**
0	60	0.504	0.533	0.03	2.509	2.608	0.099
0	90	0.579	0.566	-0.013	2.426	0.766	**-1.66**
30	0	0.582	0.582	-0.0	2.347	2.341	-0.006
30	30	0.573	0.587	0.014	1.521	2.609	1.088
30	60	0.793	0.834	0.042	1.887	2.355	0.468
30	90	0.734	0.73	-0.004	2.25	1.024	-1.226
60	0	0.597	0.597	0.0	2.304	2.302	-0.002
60	30	0.6	0.619	0.019	2.101	2.343	0.242
60	60	0.486	0.511	0.025	0.32	0.242	-0.078
60	90	0.563	0.553	**-0.01**	2.389	1.467	-0.922
90	0	0.561	0.561	-0.0	2.495	2.493	-0.002
90	30	0.523	0.532	0.009	2.436	2.385	-0.051
90	60	0.493	0.516	0.023	2.31	2.415	0.105
90	90	0.545	0.544	-0.001	2.426	2.377	-0.049

TABLE B.23: Anechoic chamber, white noise as noise, with fractional delay, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.857	0.853	-0.005	1.059	1.298	0.239
0	30	0.786	0.833	0.047	2.169	0.991	-1.178
0	60	0.71	0.745	0.035	0.963	0.492	-0.471
0	90	0.646	0.676	0.03	2.456	1.177	-1.279
30	0	0.856	0.848	-0.007	1.777	1.414	-0.363
30	30	0.79	0.823	0.033	1.466	2.412	0.946
30	60	0.701	0.777	**0.076**	0.495	0.713	0.218
30	90	0.678	0.712	0.034	0.738	1.621	0.883
60	0	0.844	0.842	-0.002	2.355	1.732	-0.623
60	30	0.792	0.829	0.037	2.603	2.381	-0.222
60	60	0.742	0.801	0.059	1.432	2.416	**0.984**
90	0	0.848	0.839	**-0.009**	2.655	0.882	**-1.773**
90	30	0.693	0.757	0.064	1.952	2.427	0.475
90	60	0.763	0.812	0.048	1.901	2.339	0.438
90	90	0.737	0.762	0.026	2.514	2.129	-0.385

TABLE B.24: Anechoic chamber, tones as noise, with fractional delay, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.857	0.603	**-0.255**	1.059	1.614	0.555
0	30	0.786	0.659	-0.127	2.169	2.076	-0.093
0	60	0.71	0.551	-0.159	0.963	2.497	1.534
0	90	0.646	0.589	-0.057	2.456	2.437	-0.019
30	0	0.856	0.6	**-0.255**	1.777	0.2	**-1.577**
30	30	0.79	0.674	-0.117	1.466	2.452	0.986
30	60	0.701	0.578	-0.123	0.495	2.48	**1.985**
30	90	0.678	0.597	-0.081	0.738	0.738	0.0
60	0	0.844	0.595	-0.25	2.355	1.84	-0.515
60	30	0.792	0.664	-0.128	2.603	1.266	-1.337
60	60	0.742	0.55	-0.192	1.432	2.224	0.792
90	0	0.848	0.594	-0.254	2.655	2.439	-0.216
90	30	0.693	0.662	**-0.031**	1.952	2.439	0.487
90	60	0.763	0.577	-0.187	1.901	2.088	0.187
90	90	0.737	0.616	-0.121	2.514	2.132	-0.382

TABLE B.25: Anechoic chamber, white noise as noise, no fractional delay, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.621	0.621	**0.0**	2.401	2.388	-0.013
0	30	0.798	0.845	0.047	2.361	1.472	-0.889
0	60	0.72	0.768	0.048	2.435	2.331	-0.104
0	90	0.697	0.712	0.015	0.632	2.365	**1.733**
30	0	0.852	0.852	**0.0**	2.499	2.506	0.007
30	30	0.802	0.845	0.043	2.158	2.329	0.171
30	60	0.563	0.599	0.036	2.131	1.699	-0.432
30	90	0.729	0.766	0.037	2.179	0.791	-1.388
60	0	0.852	0.852	**0.0**	2.289	2.289	0.0
60	30	0.795	0.84	0.044	2.383	1.551	-0.832
60	60	0.724	0.779	0.055	2.203	2.447	0.244
60	90	0.759	0.794	0.035	0.929	0.896	-0.033
90	0	0.815	0.815	**0.0**	2.35	2.351	0.001
90	30	0.785	0.847	**0.062**	2.426	2.221	-0.205
90	60	0.714	0.765	0.052	1.762	0.95	-0.812
90	90	0.783	0.806	0.023	1.825	0.407	**-1.418**

TABLE B.26: Anechoic chamber, tones as noise, no fractional delay, sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.857	0.62	**-0.238**	1.059	2.031	0.972
0	30	0.786	0.646	-0.14	2.169	2.435	0.266
0	60	0.71	0.565	-0.144	0.963	2.587	1.624
0	90	0.646	0.577	**-0.07**	2.456	2.073	-0.383
30	0	0.856	0.617	**-0.238**	1.777	3.02	1.243
30	30	0.79	0.659	-0.131	1.466	1.962	0.496
30	60	0.701	0.594	-0.107	0.495	2.423	**1.928**
30	90	0.678	0.583	-0.095	0.738	0.81	0.072
60	0	0.844	0.61	-0.234	2.355	1.29	-1.065
60	30	0.792	0.651	-0.141	2.603	1.288	-1.315
60	60	0.742	0.566	-0.176	1.432	1.591	0.159
90	0	0.848	0.61	**-0.238**	2.655	1.304	**-1.351**
90	30	0.693	0.649	-0.044	1.952	2.387	0.435
90	60	0.763	0.593	-0.17	1.901	2.424	0.523
90	90	0.737	0.601	-0.136	2.514	2.702	0.188

TABLE B.27: Anechoic chamber, white noise as noise, with fractional delay, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.621	0.626	0.005	2.401	2.415	0.014
0	30	0.798	0.846	0.048	2.361	2.369	0.008
0	60	0.72	0.767	0.047	2.435	2.347	-0.088
0	90	0.697	0.711	0.015	0.632	2.314	**1.682**
30	0	0.852	0.853	0.002	2.499	2.307	-0.192
30	30	0.802	0.845	0.043	2.158	2.551	0.393
30	60	0.563	0.604	0.041	2.131	2.259	0.128
30	90	0.729	0.767	0.037	2.179	0.874	**-1.305**
60	0	0.852	0.851	**-0.001**	2.289	2.294	0.005
60	30	0.795	0.841	0.046	2.383	2.4	0.017
60	60	0.724	0.777	0.053	2.203	2.704	0.501
60	90	0.759	0.792	0.033	0.929	2.338	1.409
90	0	0.815	0.817	0.001	2.35	2.387	0.037
90	30	0.785	0.847	**0.062**	2.426	2.292	-0.134
90	60	0.714	0.773	0.059	1.762	2.302	0.54
90	90	0.783	0.805	0.022	1.825	2.326	0.501

TABLE B.28: Anechoic chamber, tones as noise, with fractional delay, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.621	0.477	-0.144	2.401	2.244	-0.157
0	30	0.798	0.627	-0.171	2.361	2.25	-0.111
0	60	0.72	0.618	-0.102	2.435	2.426	-0.009
0	90	0.697	0.541	-0.156	0.632	2.278	1.646
30	0	0.852	0.661	-0.191	2.499	2.307	-0.192
30	30	0.802	0.603	-0.199	2.158	2.258	0.1
30	60	0.563	0.461	-0.102	2.131	2.082	-0.049
30	90	0.729	0.567	-0.162	2.179	0.823	**-1.356**
60	0	0.852	0.659	-0.193	2.289	2.299	0.01
60	30	0.795	0.61	-0.185	2.383	2.356	-0.027
60	60	0.724	0.637	**-0.087**	2.203	2.303	0.1
60	90	0.759	0.572	-0.187	0.929	2.776	**1.847**
90	0	0.815	0.662	-0.153	2.35	2.252	-0.098
90	30	0.785	0.592	-0.192	2.426	2.265	-0.161
90	60	0.714	0.589	-0.125	1.762	2.34	0.578
90	90	0.783	0.563	**-0.22**	1.825	1.423	-0.402

TABLE B.29: Anechoic chamber, white noise as noise, no fractional delay, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.621	0.621	**0.0**	2.401	2.388	-0.013
0	30	0.798	0.845	0.047	2.361	1.472	-0.889
0	60	0.72	0.768	0.048	2.435	2.331	-0.104
0	90	0.697	0.712	0.015	0.632	2.365	**1.733**
30	0	0.852	0.852	**0.0**	2.499	2.506	0.007
30	30	0.802	0.845	0.043	2.158	2.329	0.171
30	60	0.563	0.599	0.036	2.131	1.699	-0.432
30	90	0.729	0.766	0.037	2.179	0.791	**-1.388**
60	0	0.852	0.852	**0.0**	2.289	2.289	0.0
60	30	0.795	0.84	0.044	2.383	1.551	-0.832
60	60	0.724	0.779	0.055	2.203	2.447	0.244
60	90	0.759	0.794	0.035	0.929	0.896	-0.033
90	0	0.815	0.815	**0.0**	2.35	2.351	0.001
90	30	0.785	0.847	**0.062**	2.426	2.221	-0.205
90	60	0.714	0.765	0.052	1.762	0.95	-0.812
90	90	0.783	0.806	0.023	1.825	0.407	-1.418

TABLE B.30: Anechoic chamber, tones as noise, no fractional delay, sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.621	0.473	-0.148	2.401	2.251	-0.15
0	30	0.798	0.633	-0.165	2.361	2.226	-0.135
0	60	0.72	0.613	-0.107	2.435	2.333	-0.102
0	90	0.697	0.546	-0.151	0.632	2.208	**1.576**
30	0	0.852	0.656	-0.196	2.499	2.299	-0.2
30	30	0.802	0.61	-0.192	2.158	2.238	0.08
30	60	0.563	0.457	-0.105	2.131	2.116	-0.015
30	90	0.729	0.573	-0.156	2.179	1.443	-0.736
60	0	0.852	0.654	-0.198	2.289	2.302	0.013
60	30	0.795	0.617	-0.178	2.383	1.346	**-1.037**
60	60	0.724	0.632	**-0.092**	2.203	2.303	0.1
60	90	0.759	0.578	-0.181	0.929	2.322	1.393
90	0	0.815	0.657	-0.158	2.35	2.256	-0.094
90	30	0.785	0.599	-0.186	2.426	2.232	-0.194
90	60	0.714	0.584	-0.13	1.762	1.808	0.046
90	90	0.783	0.57	**-0.214**	1.825	2.321	0.496

B.3 Low-pass

TABLE B.31: Studio, white noise as noise, low-pass (4000 Hz cutoff), sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.836	0.834	-0.002	1.473	2.325	0.852
0	30	0.533	0.533	**0.0**	0.765	2.407	**1.642**
0	60	0.533	0.532	-0.001	1.191	1.559	0.368
0	90	0.542	0.54	-0.002	2.466	2.446	-0.02
30	0	0.607	0.604	-0.003	2.46	1.784	-0.676
30	30	0.561	0.56	-0.001	1.423	2.458	1.035
30	60	0.783	0.778	**-0.005**	2.488	0.385	**-2.103**
30	90	0.517	0.514	-0.002	2.537	2.306	-0.231
60	0	0.663	0.662	-0.001	2.598	2.481	-0.117
60	30	0.539	0.538	-0.002	0.829	1.972	1.143
60	60	0.562	0.559	-0.002	2.408	2.442	0.034
60	90	0.514	0.512	-0.002	2.459	2.393	-0.066
90	0	0.583	0.581	-0.002	2.475	1.513	-0.962
90	30	0.546	0.545	-0.002	1.174	2.45	1.276
90	60	0.513	0.512	-0.001	0.98	2.47	1.49
90	90	0.511	0.51	-0.0	2.454	2.464	0.01

TABLE B.32: Studio, white noise as noise, low-pass (4000 Hz cutoff), sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	Δ_{STOI}	**Unprocessed**	**Processed**	Δ_{PESQ}
0	0	0.836	0.835	-0.001	2.207	2.258	0.051
0	30	0.78	0.78	-0.001	0.988	1.369	0.381
0	60	0.504	0.503	**0.0**	2.509	2.45	-0.059
0	90	0.579	0.577	-0.002	2.426	1.319	-1.107
30	0	0.582	0.581	**0.0**	2.347	2.509	0.162
30	30	0.573	0.57	-0.003	1.521	2.058	**0.537**
30	60	0.793	0.787	**-0.006**	1.887	0.484	-1.403
30	90	0.734	0.73	-0.004	2.25	0.725	-1.525
60	0	0.597	0.596	-0.001	2.304	2.164	-0.14
60	30	0.6	0.599	**0.0**	2.101	0.509	**-1.592**
60	60	0.486	0.484	-0.002	0.32	0.303	-0.017
60	90	0.563	0.561	-0.002	2.389	2.306	-0.083
90	0	0.561	0.56	-0.001	2.495	2.427	-0.068
90	30	0.523	0.522	-0.001	2.436	1.226	-1.21
90	60	0.493	0.493	**0.0**	2.31	1.266	-1.044
90	90	0.545	0.543	-0.002	2.426	2.326	-0.1

TABLE B.33: Anechoic chamber, white noise as noise, low-pass (4000 Hz cutoff), sample rate 16 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.857	0.855	**-0.003**	1.059	1.759	0.7
0	30	0.786	0.788	**0.002**	2.169	0.404	**-1.765**
0	60	0.71	0.708	-0.002	0.963	2.43	**1.467**
0	90	0.646	0.646	0.0	2.456	2.428	-0.028
30	0	0.856	0.853	-0.002	1.777	1.838	0.061
30	30	0.79	0.789	-0.002	1.466	2.925	1.459
30	60	0.701	0.701	-0.0	0.495	0.661	0.166
30	90	0.678	0.677	-0.001	0.738	2.424	1.686
60	0	0.844	0.842	-0.002	2.355	2.436	0.081
60	30	0.792	0.79	-0.002	2.603	1.706	-0.897
60	60	0.742	0.738	**-0.003**	1.432	2.068	0.636
90	0	0.848	0.846	-0.002	2.655	1.371	-1.284
90	30	0.693	0.693	-0.0	1.952	1.269	-0.683
90	60	0.763	0.761	-0.002	1.901	2.37	0.469
90	90	0.737	0.737	0.001	2.514	2.441	-0.073

TABLE B.34: Anechoic chamber, white noise as noise, low-pass (4000 Hz cutoff), sample rate 44.1 kHz

DOA		STOI			PESQ		
$\theta(^\circ)$	$\phi(^\circ)$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{STOI}}$	**Unprocessed**	**Processed**	$\Delta_{\mathbf{PESQ}}$
0	0	0.621	0.62	-0.001	2.401	1.546	-0.855
0	30	0.798	0.797	-0.001	2.361	1.737	-0.624
0	60	0.72	0.717	-0.003	2.435	2.441	0.006
0	90	0.697	0.696	-0.001	0.632	2.054	**1.422**
30	0	0.852	0.849	-0.002	2.499	2.386	-0.113
30	30	0.802	0.801	-0.001	2.158	2.073	-0.085
30	60	0.563	0.562	-0.0	2.131	1.5	-0.631
30	90	0.729	0.728	-0.001	2.179	1.124	-1.055
60	0	0.852	0.849	-0.003	2.289	1.763	-0.526
60	30	0.795	0.795	-0.0	2.383	2.337	-0.046
60	60	0.724	0.721	**-0.003**	2.203	1.781	-0.422
60	90	0.759	0.759	0.0	0.929	0.827	-0.102
90	0	0.815	0.813	-0.003	2.35	2.193	-0.157
90	30	0.785	0.785	0.0	2.426	0.28	**-2.146**
90	60	0.714	0.711	-0.003	1.762	2.479	0.717
90	90	0.783	0.785	**0.002**	1.825	2.467	0.642

Bibliography

[1] E. C. Cherry, "Some experiments on the recognition of speech, with one and with two ears", *The Journal of the Acoustical Society of America*, vol. 25, no. 5, pp. 975–979, Sep. 1953.

[2] miniDSP. (2018). UMA-16 USB mic array, [Online]. Available: `https://www.minidsp.com/products/usb-audio-interface/uma-16-microphone-array` (visited on 04/09/2018).

[3] Python. (2018). About Python, [Online]. Available: `https://www.python.org/` (visited on 10/12/2018).

[4] Axis Communications AB. (2018). About Axis, [Online]. Available: `https://www.axis.com/en-gb/about-axis` (visited on 11/10/2018).

[5] Centers for Disease Control and Prevention. (2015). About Sound - sound frequency (pitch), [Online]. Available: `https://www.cdc.gov/ncbddd/hearingloss/sound.html` (visited on 10/12/2018).

[6] N. R. French and S. J. C. Steinberg, "Factors governing the intelligibility of speech sounds", *Journal of the Acoustical Society of America*, no. 19, pp. 90–119, 1947.

[7] J. Benesty, J. Chen, and Y. Huang, *Microphone array signal processing*. Berlin, Germany: Springer-Verlag, 2008, vol. 1, ISBN: 978-3-540-78611-5.

[8] Y. Xiang, D. Peng, and Z. Yang, *Blind source separation - dependent component analysis*. Springer, 2015, ISBN: 978-981-287-226-5.

[9] K. Qian, Y. Zhang, S. Chang, X. Yang, D. Florencio, and M. Hasegawa-Johnson, "Deep learning based speech beamforming", in *The 43rd IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP2018)*, Feb. 2018.

[10] F. A. Everest and K. C. Pohlmann, *Master handbook of acoustics*, 5th ed. New York: McGraw-Hill, 2009, ISBN: 978-0-07-160333-1.

[11] Siemens. (2018). Sound Fields: Free versus Diffuse Field, Near versus Far Field, [Online]. Available: `https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/Sound-Fields-Free-versus-Diffuse-Field-Near-versus-Far-Field/ta-p/387463` (visited on 11/12/2018).

[12] J. G. Proakis and D. K. Manolakis, *Digital signal processing*, 4th ed. United States of America: Pearson Education Limited, 2014, ISBN: 1-292-02573-5.

[13] T. I. Laakso, V. Välimäki, M. Karjalainen, and U. K. Laine, "Splitting the unit delay - tools for fractional delay filter design", *IEEE signal processing magazine*, vol. 13, no. 1, pp. 30–60, Jan. 1996.

[14] M. Swartling, "Direction of arrival estimation and localization of multiple speech sources in enclosed environments", PhD thesis, Blekinge Institute of Technology, 2012.

[15] I. McCowan, "Robust speech recognition using microphone arrays", PhD thesis, Queensland University of Technology, Australia, 2001.

[16] D. H. Johnson and D. E. Dudgeon, *Array signal processing concepts and techniques*. Prentice Hall, 1993, vol. 1, ISBN: 978-0130485137.

[17] R. G. Lorenz and S. P. Boyd, "Robust minimum variance beamforming", *IEEE transactions on signal processing*, vol. 53, no. 5, pp. 1684–1696, May 2005.

[18] J. Oh, S.-J. Kim, and K.-L. Hsiung, "A computationally efficient method for robust minimum variance beamforming", *Vehicular Technology Conference*, vol. 2, pp. 1162–1165, Jan. 2005.

[19] "IEEE Recommended Practice for Speech Quality Measurements", *IEEE Transactions on Audio and Electroacoustics*, vol. 17, no. 3, pp. 225–246, Sep. 1969.

[20] 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, Sep. 2011.

[21] International Telecommunication Union. (2019). P.862 : Perceptual evaluation of speech quality (PESQ), [Online]. Available: `https://www.itu.int/rec/T-REC-P.862/en` (visited on 01/17/2019).

[22] International Telecommunication Union (ITU). (2019). Reference implementations and conformance testing for ITU-T Recs P.862, P.862.1 and P.862.2, [Online]. Available: `https://www.itu.int/rec/T-REC-P.862-200511-I!Amd2/en` (visited on 01/17/2019).

[23] VOCAL Technologies. (2017). Blind Signal Separation (BSS), [Online]. Available: `https://www.vocal.com/blind-signal-separation/` (visited on 01/31/2019).

Printed by Tryckeriet i E-huset, Lund 2019

LUND
UNIVERSITY

Series of Master's theses
Department of Electrical and Information Technology
LU/LTH-EIT 2019-681
http://www.eit.lth.se