Acoustic Enhancement via Beamforming using Smartphones

July 20, 2015

N. van Wijngaarden 4063899

E. H. Wouters 4095421

Delft University of Technology

TU Delft Delft University of Technology

Challenge the future

PREFACE

First and foremost, our gratitude goes out to our supervisor dr. Jorge Martínez Castañeda for his excellent guidance and valuable insights throughout the entirety of this project. Without his help, we would not have been able to make such progress in a short amount of time. In addition we would like to thank prof. Bastiaan Kleijn for sharing his valuable knowledge, which helped us shape the project in an early stage. Also the interest from Jan Skoglund (Google Inc.) gave us the inspiration to put time and effort in this project to make it a success. We thank Henry den Bok for helping us to set up the measurements in the anechoic chamber at the TU Delft. Finally we want to thank Tim, Rosalie, Sjoerd and Roy for the good collaboration during this Electrical Engineering Bachelor Graduation Project.

N. van Wijngaarden 4063899
E. H. Wouters 4095421
Delft, July 20, 2015

CONTENTS

List of Figures ix

List of Tables xi

1 Introduction 3
- 1.1 Conference Calls . . . 3
- 1.2 Beamforming Algorithms . . . 3
- 1.3 Microphone Arrays . . . 4
- 1.4 Microphone Directivity . . . 4
- 1.5 Assessing Audio Quality . . . 4
- 1.6 Reading Guide . . . 4
- 1.7 Requirements . . . 4
  - 1.7.1 Requirements concerning the intended use . . . 5
  - 1.7.2 Requirements concerning the system . . . 5

2 Problem Definition 7
- 2.1 Array Signal Processing . . . 7
- 2.2 Ad-hoc . . . 7
- 2.3 Near-field . . . 7
- 2.4 Broadband . . . 7
- 2.5 Observation Model . . . 8

3 Related research 11
- 3.1 Beamforming Algorithms . . . 11
  - 3.1.1 Delay-and-sum Beamformer . . . 11
  - 3.1.2 Minimum variance distortionless response Beamformer . . . 11
  - 3.1.3 Other beamforming algorithms . . . 12
- 3.2 Evaluation of the Beamforming techniques . . . 12

4 Describing the design process 15
- 4.1 Simulation . . . 15
- 4.2 Beamforming . . . 17
  - 4.2.1 Delay-and-sum Beamformer . . . 17
  - 4.2.2 Minimum variance distortionless response Beamformer . . . 17
  - 4.2.3 MVDR with Directivity . . . 18
- 4.3 Assessing Beamforming Quality . . . 18
- 4.4 Realtime Beamforming Toolbox . . . 19
  - 4.4.1 Data Acquisition . . . 19
  - 4.4.2 Buffer . . . 19
  - 4.4.3 Synchronized Recording . . . 20
  - 4.4.4 Processing Unit . . . 20
  - 4.4.5 User Interface (UI) . . . 21

5 Results 23
- 5.1 Measurement Scenario . . . 23
  - 5.1.1 Anechoic Chamber experiments . . . 23
  - 5.1.2 Office room experiments . . . 25
  - 5.1.3 Simulations of the office room experiments . . . 25
  - 5.1.4 Results of the office room experiment . . . 26

6 Discussion **29**
6.1 Performance of the Beamforming Algorithms . . . 29
6.2 Possible improvements . . . 29
6.2.1 Localization . . . 29
6.2.2 Distributed algorithm . . . 30
6.2.3 Realtime Beamforming Toolbox . . . 30

7 Conclusion **31**
7.1 Simulation . . . 31
7.2 Measurements . . . 31
7.3 Location errors . . . 31
7.4 Directivity . . . 32

A Ethical considerations **33**
A.1 The need for ethical discussion . . . 33
A.2 Unintentional use cases . . . 33
A.3 Beamformer contribution . . . 33
A.4 Applications . . . 33
A.5 Final remarks . . . 34

B Detailed Results **35**
B.1 Simulation . . . 35
B.1.1 Simulation 1 . . . 36
B.1.2 Simulation 2 . . . 37
B.1.3 Simulation 3 . . . 38
B.1.4 Simulation 4 . . . 39
B.1.5 Simulation 6 . . . 40
B.2 Office room experiments . . . 41
B.2.1 Office room experiment 1 . . . 41
B.2.2 Office room experiment 2 . . . 42
B.2.3 Office room experiment 3 . . . 43
B.2.4 Office room experiment 4 . . . 44
B.2.5 Office room experiment 5 . . . 45
B.3 Anechoic chamber experiments . . . 46
B.3.1 Anechoic chamber experiment 1 . . . 46

Bibliography **47**

NOMENCLATURE

$\hat{\mathbf{w}}$	Vector of estimated complex weights
$\hat{S}(\omega)$	Estimated source signal
$\hat{x}(n)$	Distorted speech signal in the time domain
λ	Wavelength [m]
ω	Frequency [rad/s]
$\overline{x}(n)$	Average estimated clean speech signal in the time domain
Φ	Covariance matrix of the microhone signals
ϕ	Elevation [°]
τ	Delay of a signal [s]
$\mathbf{d}(\omega,\theta,\phi)$	Look vector of the array
$\mathbf{p}$	Position vector in space [m]
$\mathbf{w}(\omega)$	Vector of complex weighing factors
$\mathrm{SNR}_{\mathrm{seg}}$	Segmental signal-to-noise ratio [dB]
θ	Azimuth [°]
a	Attenuation coefficient [m^{-1}]
$A(\omega,\theta,\phi)$	Attenuation coefficient in the frequency domain, as a function of the look direction
c	Speed of sound [m/s]
$d(t,\theta,\phi)$	Array response vector [s]
$G(\omega,\theta,\phi)$	Directivity pattern of a microphone in the frequency domain
$g(t,\theta,\phi)$	Directivity pattern of a microphone in the time domain
G_{A}	Array-gain
$H(\omega)$	Acoustic transfer function in the frequency domain
$h(t)$	Acoustic transfer function in the time domain
L	Largest array dimension [m]
L_f	Frame length of the segmental signal-to-noise ratio [s]
M	Number of microphones
N	Number of samples
r	Distance between source and mic [m]
$S(\omega)$	Source signal in the frequency domain
$s(t)$	Acoustic source signal in the time domain

SNR	Signal-to-noise ratio [dB]
t	Time [s]
$V(\omega)$	Undesired signal components in the frequency domain
$v(t)$	Undesired signal components in the time domain
$x(n)$	Clean speech signal in the time domain
$Y(\omega)$	Received audio signal by a microphone in the frequency domain
$y(t)$	Received audio signal by a microphone in the time domain
MOS	Mean opinion scores
PESQ	Perceptual evaluation of speech
STOI	Short-Time Objective Intelligibility measure

LIST OF FIGURES

1.1 An example usage scenario of the system in a conference call . . . 3

2.1 Room representation for the observational model [1] . . . 9

3.1 Block diagram of the eMVDR [1] . . . 12

4.1 Image method for simulating a rooms [2] . . . 16
4.2 Simulated Impulse response without reverberations . . . 16
4.3 Simulated Impulse response with reverberations . . . 16
4.4 General overview of the PESQ algorithm [3] . . . 19
4.5 Beamforming Toolbox flowchart . . . 20
4.6 User interface implemented in MATLAB . . . 21

5.1 Setup of the measurements in the anechoic room . . . 23
5.2 Audio signals and spectrogram anechoic chamber experiment 1: With an interfering audio source 24
5.3 Intelligibility anechoic chamber experiment 1: With an interfering audio source . . . 24
5.4 System configuration during the office room experiments . . . 25

B.1 Audio signals simulation 1 . . . 36
B.2 Intelligibility simulation 1 . . . 36
B.3 Audio signals simulation 2 . . . 37
B.4 Intelligibility simulation 2 . . . 37
B.5 Audio signals simulation 3 . . . 38
B.6 Intelligibility simulation 3 . . . 38
B.7 Audio signals simulation 4 . . . 39
B.8 Intelligibility simulation 4 . . . 39
B.9 Audio signals simulation 6 . . . 40
B.10 Intelligibility simulation 6 . . . 40
B.11 Audio signals office room experiment 1 . . . 41
B.12 Intelligibility office room experiment 1 . . . 41
B.13 Audio signals office room experiment 2 . . . 42
B.14 Intelligibility office room experiment 2 . . . 42
B.15 Audio signals office room experiment 3 . . . 43
B.16 Intelligibility office room experiment 3 . . . 43
B.17 Audio signals office room experiment 4 . . . 44
B.18 Intelligibility office room experiment 4 . . . 44
B.19 Audio signals office room experiment 5 . . . 45
B.20 Intelligibility office room experiment 5 . . . 45
B.21 Audio signals anechoic chamber experiment 1 . . . 46
B.22 Intelligibility anechoic chamber experiment 1 . . . 46

LIST OF TABLES

5.1 Experimental results of the anechoic chamber measurement 25
5.2 The simulated scenarios 26
5.3 Simulated results 27
5.4 Experimental results 28

B.1 The simulated scenarios 35

ABSTRACT

This thesis presents a comparison between two different beamforming algorithms, the delay-and-sum beamformer (DSB) and the minimum variance distortionless response (MVDR) beamformer using both objective and subjective metrics. Simulations and measurements have been performed on near-field scenarios with an ad-hoc microphone array. Microphones from multiple smartphones are used to record the audio. Smartphone microphones are designed to have a distinct directivity gain pattern. In this research, the MVDR algorithm is extended to compensate for these directivity responses. In simulated scenarios, without the directivities of the microphones, the average improvement in segmental signal to noise ratio is $5.3dB$ for DSB and $11.9dB$ for MVDR. In all the simulated scenarios, the MVDR beamforming algorithm outperforms the DSB algorithm. Results from measurements in an office room scenario only showed an improvement when applying the DSB algorithm. The results from measurements in the anechoic chamber at the TU Delft show $1.9dB$ improvement for DSB and $2.8dB$ for MVDR. Compensating for the directivity in the implemented MVDR mainly added noise to the signal.

1

INTRODUCTION

Companies and many other organizations depend more and more on teleconferencing as a means of communications. This is a viable solution that can reduce both cost and environmental impact of traveling long distances for meetings. Commercially available, state of the art teleconferencing products usually require a controlled acoustic environment equipped with expensive hardware [4]. However, nowadays it becomes increasingly more common to possess a smartphone. This brings forward the opportunity to use the microphones from multiple smartphones for audio conferencing. There are multiple aspects of audio quality to be considered in developing and designing a conference telephone system. A few examples are noise and reverberations [5]. These must be filtered from the microphone signals to avoid noise and echoes [6].

1.1. CONFERENCE CALLS

In a conference call, the interesting audio source is usually one person talking in a noisy room full of people. Therefore it is a challenge to enhance the sound from the source while suppressing the surrounding (noise) sources. Beamforming provides an effective and versatile means of spatial filtering [5]. This is useful when distinguishing desired signal components from noise and interference based on their locations.

1.2. BEAMFORMING ALGORITHMS

In this research, the delay-and sum beamformer and the minimum variance distortionless response (MVDR) beamformer are investigated. The MVDR beamformer is a very widely studied and used beamforming algorithm [7], and is also used in commercially available arrays [4]. The MVDR beamformer estimates the desired signal while minimizing the variance of the noise component of the formed estimate. The locations of the mi-

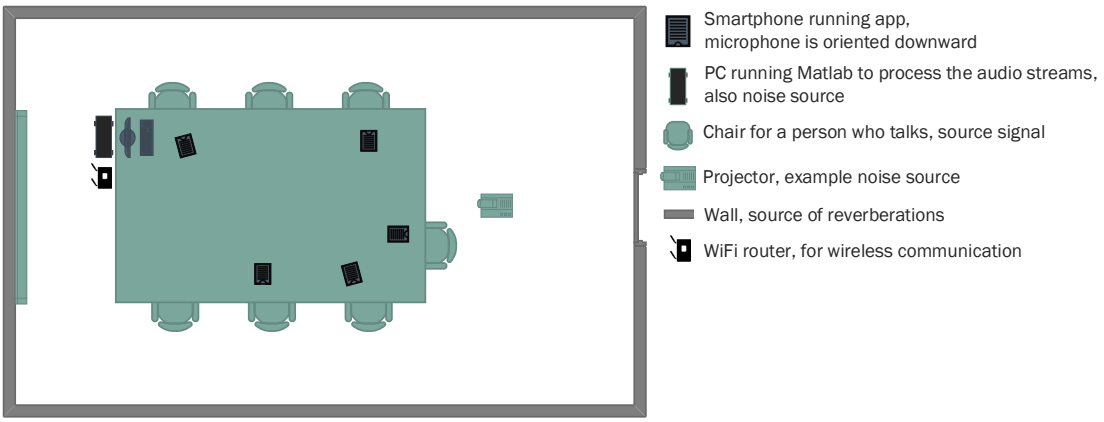


Figure 1.1: An example usage scenario of the system in a conference call

crophones are used to determine the direction of arrival (DOA). There are existing algorithms which can do source localization by using spectral estimation concepts and time-difference of arrival (TDOA) information [8]. Inaccuracies in the locations of the microphones degrade the performance of the MVDR beamformer significantly [5, 9]. Fortunately the microphone and source locations are known a priori in many cases [10]. For this research perfect sound source localization (SSL) is assumed for both microphones and sources.

1.3. MICROPHONE ARRAYS

Microphone arrays are a common way to improve the sound quality [8]. The diversity in the received signals is exploited by setting different gains to each microphone, depending on the location of the source and the interference. This is generally referred to as beamforming. Early designs were generally "fixed" beamformers like the delay-and-sum beamforming [6], adapting only to the location of the desired source. State of the art beamformers use statistical approaches which, in addition to the locations, also depend on assumptions about the acoustic composition of the signals [11]. Iterative methods are also available which can adapt to specific noise and acoustic conditions [1, 12]. The latter methods have a considerably higher complexity [13]. A lot of research has been done into a class of algorithms known as robust MVDR beamformers [1, 14, 15]. These algorithms improve upon previous work on two fronts, by extending the region where the source can be located and thus lowering the susceptibility to errors in the localization, and by including nonuniformity in the microphones directional response, known as microphone directivity.

1.4. MICROPHONE DIRECTIVITY

Real microphones may have distinct, directional responses. Experiments done by Ba et al. [1] show that the traditional MVDR beamforming algorithm and other existing algorithms perform well when omnidirectional microphones are used, but do not provide much enhancement when directional microphones are used. Simulations done by Martínez et al. [11] show improvements of up to $6.1dB$ in the segmental signal to noise ratio when using the measured directivity compared to assuming the microphones are omnidirectional [15]. Aforementioned work used low cost smartphone microphones and the directivity pattern was measured in two dimensions. We will show in chapter 5 that including the directivity impulse response produced erratic results. For the smartphones used in this work the directivities have been measured by our fellow team members [16]. This result is incorporated in the robust MVDR algorithm in section 4. This has not been exploited much in the existing literature and in this thesis some of the challenges of implementing a realtime robust MVDR beamforming algorithm which adopts microphone directivity have been explored.

1.5. ASSESSING AUDIO QUALITY

In order to compare different beamformers, parameters and smartphone configurations an attempt has been made to write a realtime audio beamforming toolbox for MATLAB which can interface with an Android app developed by our fellow team members [17]. With this toolbox the impact of different aspects of beamforming on the quality estimators can be visualized in realtime. This enables rapid testing of different parameters and configurations. There are different quality estimators to analyze the performance of a beamformer. There are power related measures and intelligibility measures. Intelligibility is the degree to which speech can be understood and hence the corresponding measures can be evaluated to determine the effect of the beamformers on the speech intelligibility of the signals [18].

1.6. READING GUIDE

This thesis is organized as follows. The last section (1.7) of this introduction contains a schedule of requirements which was made at the start of the project. In chapter 2, the problem definition and observation model of this research will be defined. The research that has already been conducted in this topic will be discussed in section 3. The design process is described in section 4. The results are presented in section 5. A discussion on these results is given in section 6. A conclusion is drawn in section 7. Detailed results from the data analysis can be found in appendix B.

1.7. REQUIREMENTS

The aim of the system that will be implemented is to record audio, using smartphones, in a room with multiple audio sources and compare the performance of different beamforming algorithms. This comparison is

done by showing different quantitative analyses related to a beamforming algorithm in a way that a user can compare the differences at a glance.

1.7.1. REQUIREMENTS CONCERNING THE INTENDED USE

1. The system must have the possibility to either select local audio files for simulation or post-processing, or connect to smartphones for realtime calculations.
2. The noise has to be suppressed while maintaining the sound level of the source.
3. The noise source can be located anywhere, except at the source location.
4. The system must be able to switch between different types of beamformers.
5. The system must be able to compare the received audio from one of the smartphones before and after applying one of the beamformers.
6. The system must calculate and compare intelligibility metrics to determine the quality of the output for speech purposes.
7. The system must calculate and compare power related metrics to determine the quality of the output.
8. The system must be able record sound using smartphones.
9. The system must be able to play the beamformed audio over the speakers.
10. The system must be able to connect to the smartphones and handle the streaming data.
11. All of the requirements should be achieved in situations where the near-field assumptions hold.

1.7.2. REQUIREMENTS CONCERNING THE SYSTEM

1. The beamforming algorithms can be applied to real-time audio signals or previously recorded files.
2. For realtime beamforming, the communications between matlab and the smartphones will be monitored. For each smartphone individually the location, orientation and if the smartphone is placed face up or face down on the table can be entered.
3. Recording can be started and stopped by pressing a start/stop button.
4. The different beamformers can be turned on or of by ticking the corresponding box.
5. Each of the different kinds of intelligibility and power related calculations can be selected or deselected.
6. While recording, the location(s) of the source(s) and microphone(s) will be plotted on the screen in 3D.
7. After recording, the relevant data can be saved to an audio file, including sufficient data labels such as locations to make post-processing possible.
8. In the non realtime case, the user will be able to select a time frame which will be beamformed and over which the intelligibility and power related measures will be computed.

2

PROBLEM DEFINITION

2.1. ARRAY SIGNAL PROCESSING

In general array signal processing is concerned with three tasks: detection of noise signals, estimation of the direction of arrival and beamforming. The locations of the microphones and speakers in the room are usually stationary. These locations hold valuable information which can be used when processing the audio signal. An elaborate method to process this signal is beamforming [19]. This is a statistical method to estimate parameters, which is further complicated by the spatial dimensions of the array. Krim and Viberg formulate the problem of sensor array processing as: *"... the estimation of parameters by fusing temporal and spatial information, captured via sampling a wavefield with a set of judiciously placed antenna sensors."* [20].

2.2. AD-HOC

The techniques used in beamforming are not strictly valid for microphone sensor arrays, but may also hold for other types of sensor arrays. The geometry, size and number of the sensor array can vary. The geometry of the sensor array may vary from linear, circular or spherical to arbitrary geometries. This research focuses on the last type of geometry, which we will call an *ad-hoc microphone array*.

2.3. NEAR-FIELD

There is an important distinction to be made with respect to the dimensions of the array or *aperture* size. There exist different methods for *far-field* and *near-field* beamforming. In far-field one can assume that the wave field originates from a distance far greater than the dimensions of the array, which leads to the assumption that the wave field has a flat wave front. This means that the Direction Of Arrival (DOA) is the same for every sensor in the array. In near-field beamforming the distance from the source of the wave field to the array is comparable in magnitude to the dimensions of the array. This means that every sensor gets a different DOA and amplitude of the signal.
An often used rule to decide whether the far-field approximation is valid is $r = 2L^2/\lambda$ where r the distance from the microphone array to the source, L the largest array dimension and λ the operating wavelength [21]. In the scenario outlined in section 1 the dimensions of the microphone array are large with respect to the distance between the array and the source. Therefore, according to the earlier mentioned rule, we can assume that the sources of interest are in the near-field region [11]. In addition, the positioning of the microphones is arbitrary, but known. As such, the main focus for our implementation will be near-field beamforming with ad-hoc microphone arrays.

2.4. BROADBAND

There is one more distinction to be made in type of the beamformer. There are different design techniques for broadband and narrowband beamformers. In speech communication applications broadband design implies design over several octaves. In order to use the processed signals for speech communications, the beamformers have to perform well over the frequency band from $300Hz$ to $3400Hz$ [22].

2.5. OBSERVATION MODEL

Figure 2.1 shows a representation of a room with some x, y and z dimensions and M microphones. An acoustic source signal $s(t)$ originating from a position $\mathbf{p}^j_{src}$ will propagate through the room and reach all the microphones positioned at $\mathbf{p}^i_{mic}$. At a time t the sampled microphone signal for each microphone consists of delayed and attenuated versions of the source signal $a_i s(t-\tau_i)$ with undesired signal components $v_i(t)$ added to it. The acoustic transfer function (ATF) of the room is denoted by $h_i(t)$. For this research the ATF of an anechoic room is assumed unless otherwise specified:

$$H_{i,j}(\omega) = \frac{\exp(-j\omega\frac{r_{i,j}}{c})}{4\pi r_{i,j}} \tag{2.1}$$

with

$$r_{i,j} = \left\| \mathbf{p}^i_{mic} - \mathbf{p}^j_{src} \right\| \tag{2.2}$$

where c is the speed of sound; $\mathbf{p}$ is the position vector in space. In many cases there will be only one source of interest, so the j index may be omitted. The ATF is represented in the Fourier domain. The positions and orientations of the microphones and sources are assumed to be known. The ATF describes the set of all the propagation paths from a source to a microphone and hence the convolution between the sound coming from the source and the impulse response results in the sound recorded by the microphones. If the microphone has nonuniform directivity gain the room ATF is convolved over time with the directivity pattern of the microphone $g(t,\theta,\phi)$:

$$d_i(t,\theta,\phi) = h_i(t) * g_i(t,\theta,\phi) \tag{2.3}$$

where "$*$" denotes the convolution; $g_i(t,\theta,\phi)$ is the directivity impulse response [16] of the specific microphone, θ and ϕ are the azimuth and elevation angles between the source and the respective microphone. This directivity pattern was measured in the anechoic chamber of the Faculty of Applied Sciences at TU Delft for the smartphones used in this research. For omnidirectional microphones a Kronecker delta can be used for g_i, or it can simply be omitted. We can combine the attenuation and delay for all microphones into one vector:

$$\mathbf{d}(\omega,\theta,\phi) = [A_1(\omega,\theta,\phi)e^{-j\omega\tau_1}, A_2(\omega,\theta,\phi)e^{-j\omega\tau_2}, ..., A_M(\omega,\theta,\phi)e^{-j\omega\tau_M}]^T \tag{2.4}$$

where $\mathbf{a}^T$ denotes the transpose of the vector $\mathbf{a}$. The vector $\mathbf{d}$ is called the *look vector* of the array. A_i is the combined attenuation factor due to the ATF and the directivity τ_i is the combined delay due to the ATF and the directivity.
We model the received signals $y_i(t), i \in \{1,...,M\}$ at each sensor using equation 2.3 as:

$$y_i(t) = a_i s(t-\tau_i) * g_i(t,\theta,\phi) + v_i(t) \tag{2.5}$$

where $v_i(t)$ denotes the undesired signal components which consist of noise and reverberations (the dotted lines in figure 2.1) and signals from interfering sources. a_i is the attenuation factor due to the ATF. The delay between when the signal is at the reference microphone and the i^{th} microphone is $\tau_i = \left\| \mathbf{p}^i_{mic} - \mathbf{p}_{src} \right\| / c$ where c denotes the speed of sound.

In the Fourier domain the received signal described in equation 2.5 for microphone i can be written as:

$$Y_i(\omega) = A_i(\omega)S(\omega)e^{-j\omega\tau_i}G_i(\omega,\theta,\phi) + V_i(\omega) \tag{2.6}$$

The attenuation factor $A_i(\omega)$ can depend on the frequency. $G_i(\omega,\theta,\phi)$ is the directivity pattern of the microphone in the Fourier domain, for nonuniform directivity microphones on the respective azimuth and elevation; $V_i(\omega)$ is the additive noise in the Fourier domain.

By choosing the values for $A_i(\omega,\theta,\phi)$ and τ_i one can maximize the gain (or steer the beam) in any direction spanned by the position vectors of the array. The desired signal can be recovered by multiplying the microphone signals by a complex weighing factor $\mathbf{w}$:

$$\hat{S}(\omega) = \mathbf{w}^H(\omega)\mathbf{Y}(\omega) \tag{2.7}$$

where $\mathbf{a}^H$ denotes the Hermitian transpose of the vector $\mathbf{a}$. The estimated signal $\hat{S}(\omega)$ can finally be transferred back to the time domain to arrive at the beamformed audio signal. Sections 4.2.1 and 4.2.2 go into how to calculate the weights $\mathbf{w}(\omega)$.

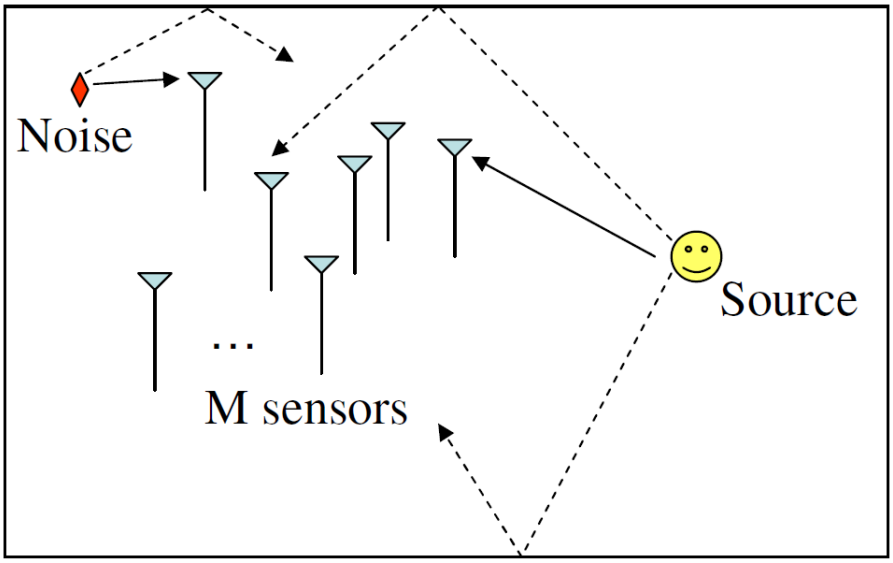


Figure 2.1: A source incident on an array of M sensors in the presence of noise and reverberations [1]

3

RELATED RESEARCH

With the increase of smartphone usage it is no surprise that a lot of research has been done considering beamforming algorithms. Many different algorithms have been proposed to use to perform beamforming with [5, 12, 20]. In this chapter, related research relevant to this topic will be discussed.

First common methods and the state of the art in beamforming research for acoustic and speech enhancement will be discussed. Then the different measures to evaluate the performance of these algorithms are introduced and to conclude ways to implement localization on the smartphones and distributed computation of the beamforming algorithm will be looked into.

3.1. BEAMFORMING ALGORITHMS

Microphone arrays are a common way to improve sound quality; the diversity in the received signals is exploited by setting different gains to each microphone, depending on the location of the source and on the auxiliary sources and noise. This is generally referred to as beamforming. There are different beamforming algorithms each with their own advantages and disadvantages. The delay-and-sum beamformer (DSB) is one of the earliest beamforming algorithms. It is a "fixed" beamformer, meaning that it only adapts to the locations of the microphones and the desired source. It delays and optionally weighs the microphone signals based on the distances and the speed of sound and the attenuation of the soundwave respectively before taking the average of the signals. The MVDR beamformer is one of the most widely used beamformers and can be used for both speech dereverberation and noise reduction [6].

3.1.1. DELAY-AND-SUM BEAMFORMER

The delay-and-sum beamformer (DSB) is an important approach to dereverberation and provides a reference to compare other methods with. The expected improvement in the Direct to Reverberant Ratio (DRR) of the DSB only depends on the distance between the source and the array and the seperation of the microphones and does not depend on reverberation time, which is the time it takes for the reverberant energy to drop to 60 dB below the original level. When the microphones are seperated by a half wavelength, perfect dereverberation is achieved [6]. For spatially white noise only, the DSB is the optimal beamformer [8].

White noise in general is a random signal with a constant power spectral density. It has a constant bandwidth and does not depend on the frequency [23]. Spatially white noise means that the noise is uncorrelated amongst all sensors in the array [20].

3.1.2. MINIMUM VARIANCE DISTORTIONLESS RESPONSE BEAMFORMER

The minimum variance distortionless response (MVDR) beamformer [6] is one of the most widely used beamformers. This beamforming algorithm can suppress coherent noise sources well. However, in a spatially white noise field, the MVDR beamformer reduces to a DSB [6]. The MVDR is also excessively sensitive to source location and microphone gains [1].

There are multiple proposed variations on the standard MVDR beamformer. The first variation on the steered beamformer based locators with an increased accuracy in reverberate rooms is proposed by Ba et al. [1]. This algorithm is specifically designed for directional microphones in a teleconference setting. The algorithm uses voice activity detection to differentiate between speech and noise in the sound at the input

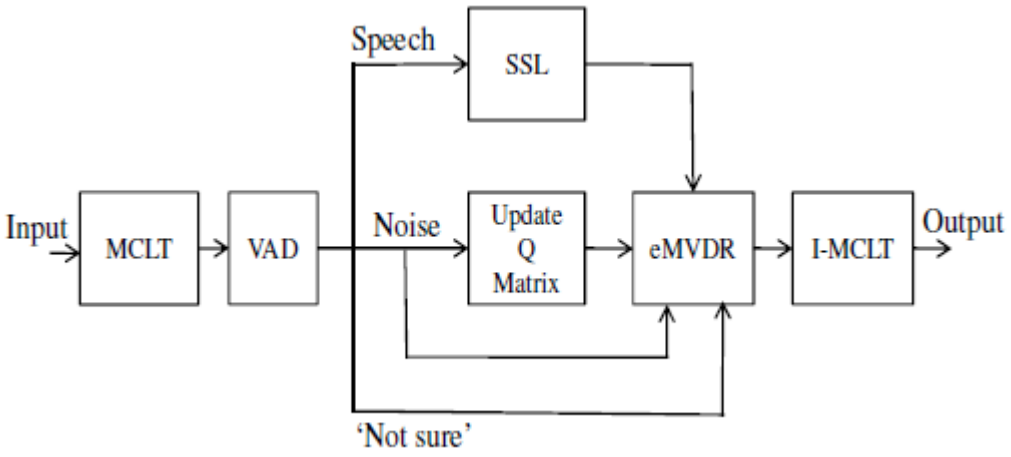


Figure 3.1: Block diagram of the eMVDR [1]

of the beamformer. The noise component of the input will be used to update the covariant matrix of the beamformer. The block diagram of this algorithm can be found in figure 3.1.

In the research by Ba et al. [1] a comparison is made between an MVDR beamformer with and without gain compensation for the microphone directivity. However, only circular microphone arrays are considered in this paper. The proposed eMVDR, which stands for enhanced MVDR, algorithm with the directivities included, outperforms the traditional MVDR in all ten of the experimented scenarios. A different variation on the MVDR, which includes directivities in the beamformer, is proposed by Gaubitch et al. [11]. This research also concludes that including the directivities in the beamformer significantly increases the performance of the beamformer.

3.1.3. OTHER BEAMFORMING ALGORITHMS

Another beamforming algorithm, which is proposed by Martinez et al. [15], uses an eigenfilter structure to focus on regions instead of just an angle. The consequence of this variation is an increase in the robustness of the beamformer.

Another kind of beamformer is the generalized sidelobe canceller (GSC) [12]. One of the disadvantages of this beamformer is the robustness problem [8]. The beamformer suffers from target-signal cancellations due to errors in the steering vector. The generalized sidelobe canceller computes the weight factors in a different way than the MVDR beamformer resulting in a much lower complexity.

3.2. EVALUATION OF THE BEAMFORMING TECHNIQUES

To investigate the performance of our algorithms, the results will be evaluated using different intelligibility measures [24, 25]. The measures used to evaluate a beamformer are power related measures and intelligibility measures. The power related measure show a ratio between powers, for example the SNR shows the ratio of signal to noise. This can be used to determine if a beamforming algorithm suppresses the noise while keeping the signal power the same. Intelligibility measures determine the extend to which a signal is understandable for speech applications. In this thesis the beamforming algorithms are used to enhance speech, which is why the intelligibility measures are interesting to investigate.

Different power related measures exist which can be used to investigate the performance of a beamformer. A few of them are:

- Global SNR[dB]
- Segmental SNR [dB]
- White noise gain
- Array-gain

The global and segmental SNR are calculated by determining the ratio between the power of the signal and the power of the noise. The global SNR uses the complete signal to determine this ratio. The segmental SNR calculates the average SNR of only those frames of the sound signal which exhibit an SNR larger than $0dB$ [8]. The segmental SNR measures the amplitude distortion of the sound signal and will therefore be well correlated with perceived distortions.

The shows the ability of a beamformer to suppress spatially uncorrelated noise [8]. White noise gain is defined as the output power due to unit variance white noise in the sensor [5]. The white noise gain is represented by $\mathbf{w}^H\mathbf{w}$. If the white noise gain is high, the beamformer output will have a poor SNR due to white noise contributions.

The array-gain shows the improvement of the signal-to-noise ratio between one of the microphones and enhanced output of the array and can be calculated as follows:

$$G_A = \frac{\text{SNR}_{\text{array}}}{\text{SNR}_{\text{microphone}}} \tag{3.1}$$

An intelligibility measure that shows a good correlation with speech intelligibility is the Short-Time Objective Intelligibility measure (STOI) [24, 26]. STOI is based on the correlation between short-time segments of the clean and processed signal. The short-time segments used were approximately $400ms$. This algorithms was specifically designed for a sample-rate of $10000Hz$ to cover the relevant frequency range for speech. The STOI is expressed in the range 0 to 1 where a higher value means that the signal is more intelligible.

Another widely used intelligibility measure is Perceptual Evaluation of Speech Quality (PESQ). PESQ predicts the perceived quality that would be given to the processed signal by subjective listening tests. This measure shows a high correlation with the subjective Mean Opinion Scores (MOS) [3, 25].

The mean opinion score is a test that results in an indication of the quality of a phone call on a scale from 1 to 5, where 1 stands for bad and 5 stands for excellent. There are guidelines, given by ITU-T recommendation *P.*800 [27], for the environment of the speaker that should be followed when conducting such a test.

It is expected that an improvement in SNR results in an improvement in MOS [28]. This means that the PESQ should also improve since MOS and PESQ show a high correlation.

4

DESCRIBING THE DESIGN PROCESS

Beamformers are useful when distinguishing desired signals from noise and interference based on their locations [10]. There are different beamforming algorithms with their own advantages and disadvantages.

In this research, the delay-and-sum beamforming algorithm [29] and the MVDR algorithm [7] will be discussed. The DSB will be considered since this is a good baseline method to compare the other beamformer with. The MVDR has been chosen since the generalized sidelobe cancellar beamformer suffers from steering vector errors and the MVDR suppresses coherent noise well.

4.1. SIMULATION

In this research, a simulation is used to simulate the effect of a reverberant room on the acoustic signals received at multiple microphones. This algorithm has been implemented in MATLAB to simulate recording multiple audio sources in a room with multiple microphones. The simulation will be used to compare to the experimental results. The acoustic impulse response of a room characterizes the acoustics of that room. The convolution between the sound coming from the source and the impulse response results in the sound recorded by the microphones. The simulated microphones are omnidirectional, which means that the directivities of the microphones are not included in this simulation.

The parameters that can be set for this simulation are:

- Locations of the audio sources and the microphones.
- Size of the room.
- Reflection coefficients.
- SNR at which Gaussian white noise will be added.
- Amount of random errors added to the microphone positions.
- Reverberations on or off.

Local stored audio files are expected as source files. The audio files will be trimmed down or zero padded to match the simulation time. Then the implemented algorithm calculates the impulse response of a room. The locations of the sources and the locations of the microphones and the size of the room will be used to calculate the impulse response between each source and each microphone. The impulse response is calculated using the image method [2]. This method mirrors the walls of a room and puts the sources in imaginary rooms next to the room in which the source and receiver are placed. This concept is illustrated in figure 4.1. The standard image method has been modified to simulate received echo arrival time to estimate the intermicrophone phase [30].

After calculating the impulse response of the room it is possible to determine the signal recorded by each microphone. The convolution of the source signals with the impulse responses result in an output signal for each microphone.

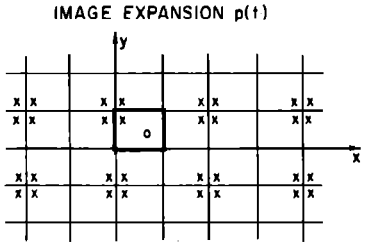


Figure 4.1: Image method, "x" stands for an audio source and an "o" stands for a microphone. [2]

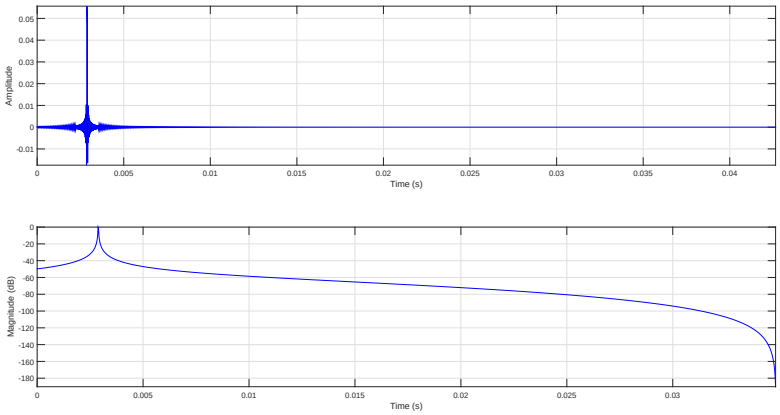


Figure 4.2: Simulated impulse response without reverberations in time-domain

This is done by calculating the convolution between the impulse response of the room and the associated audio source. After this is done for each audio source and each microphone, white noise is added to make the simulation more realistic.

Having implemented this algorithm, the choice can be made to either record audio and apply the beamforming algorithm to that or use the simulation to create the input for the beamforming algorithm. The simulation calculates the acoustic impulse of the room which will be used to generate the sound recorded by the microphones. One of the impulse responses for a simulation without reverberations and one impulse response for a simulation with reverberations is plotted and can be found in figure 4.2 and 4.3.

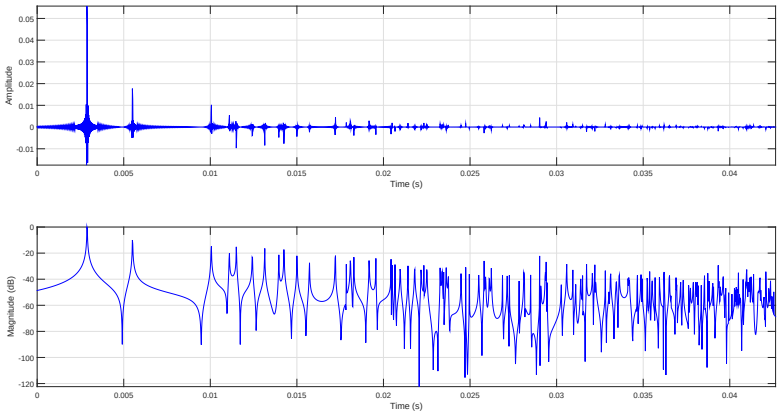


Figure 4.3: Simulated impulse response with reverberations in time-domain

4.2. BEAMFORMING

The technique called beamforming combines sensor array signals in some optimal manner such that the output approximates the desired source and suppresses the undesired noise/interferers. By choosing sensible weights a beamformer can perform spatial filtering with the goal of maximizing the intelligibility of the received signal.

The goal of beamforming is to estimate the desired signal $\hat{S}$ from equation 2.7 as a linear combination of the data collected at the array. In other words, we would like to determine an $M \times 1$ vector weights $w(\omega)$ such that $\mathbf{w}^H(\omega)\mathbf{Y}(\omega)$ is a good estimate of $S(\omega)$.

The delay-and-sum beamformer (DSB) is one of the earliest beamforming algorithms. It is a "fixed" beamformer, meaning that it only adapts to the locations of the microphones and the desired source. The MVDR beamformer is one of the most widely used beamformers and can be used for both speech dereverberation and noise reduction.

4.2.1. DELAY-AND-SUM BEAMFORMER

The delay-and-sum beamformer (DSB) first delays and weighs the microphone signals and then sums or averages the results for all the microphones. For the simple DSB the complex weights $\mathbf{w}$ usually consists of only the delays. In near-field (Section 2.3 applications the DSB may also perform inverse distance weighing in addition to delaying the signal to correct for attenuation losses. In the time domain the output $\overline{x}[n]$ can be written as [6]:

$$\overline{x}[n] = \sum_{i=1}^{M} w_i x_i[n - \tau_i] \tag{4.1}$$

If we take the discrete time Fourier transform and move the delay into the weight we get for the weights $\mathbf{w} = [w_1, w_2, \ldots, w_i, \ldots, w_M]$:

$$w_i(\omega) = \frac{1}{M} e^{-j\omega\tau_i} \tag{4.2}$$

or with inverse distance weighing:

$$w_i(\omega) = \frac{1}{4\pi r_i M} e^{-j\omega\tau_i} \tag{4.3}$$

with M the number of microphones, $i \in \{1, 2, \ldots, M\}$, w_i the weight applied to the i^{th} microphone and τ_i the propagation delay in samples. This weight can be used to estimate the source signal described in equation 2.7.

The best asset of the DSB is that the direct path components of the microphone signals will be coherent and therefore added constructively. The incoherent components, due to reverberation and noise, will be attenuated.

4.2.2. MINIMUM VARIANCE DISTORTIONLESS RESPONSE BEAMFORMER

A general description of MVDR beamformers in room acoustics and the trade-off between speech dereverberation and noise reduction is described by Cohen et al.[7] and Habets et al. [31]. The MVDR beamformer is derived by minimizing the energy of the observed signal, subject to the constraint that the signal in the desired look position remains undistorted [5]. It minimizes the variance of the noise component of $\mathbf{w}^H\mathbf{Y}$ from equation 2.7, subject to a constraint of gain 1 in the look direction [1]. The corresponding weight vector $\mathbf{w}$ is the solution to the following optimization problem:

$$\hat{\mathbf{w}} = \underset{\mathbf{w}}{\operatorname{argmin}} \mathbf{w}^H \Phi_{\mathbf{Y}} \mathbf{w} \quad \text{s.t.} \quad \mathbf{w}^H \mathbf{d} = 1 \tag{4.4}$$

where $\Phi_{\mathbf{Y}}$ is the covariance matrix of the observed signal[1]. $\Phi_{\mathbf{Y}}$ is given by $\Phi_{\mathbf{Y}} = E[\mathbf{Y}\mathbf{Y}^H]$, where $E[\bullet]$ denotes the expected value function. This optimization problem has an elegant closed-form solution published by Cox et al.[33]:

$$\hat{\mathbf{w}} = \frac{\Phi_{\mathbf{Y}}^{-1}\mathbf{d}}{\mathbf{d}^H \Phi_{\mathbf{Y}}^{-1}\mathbf{d}} \tag{4.5}$$

[1] This is actually true for Minimum Power Distortionless Response (MPDR), for traditional MVDR the covariance matrix of the combined noise signals sould be used, e.g. the covariance matrix of the reflected paths and auxilary sources [32].

The denominator of this equation contains a normalization factor which enforces the gain constraint in the look direction. The numerator contains the minimized energy information for MPDR, or minimized noise information for traditional MVDR, for the respective microphones.

4.2.3. MVDR WITH DIRECTIVITY

To illustrate how to include the microphone directivities in the algorithm a similar approach to Gaubitch et al. [11] will be followed, which is outlined in this section. The theory of how to include directive (non-omnidirectional) microphones into the MVDR algorithm was briefly covered in section 2.5. The difference is not in the algorithm, but in the used impuls response.
For omnidirectional microphones the signal received at each microphone (Y_i) consists of the following components:

$$Y_i(\omega) = S(\omega)H_i(\omega) + V_i(\omega) \tag{4.6}$$

When the microphones are directive the signal is convolved with the microphone directivity $G(\omega,\theta,\phi)$. We end up with:

$$Y_i(\omega) = S(\omega)H_i(\omega)G_i(\omega,\theta,\phi) + V_i(\omega) \tag{4.7}$$

The estimated source signal is again calculated by:

$$\hat{S}(\omega) = \mathbf{w}^H(\omega)\mathbf{Y}(\omega) \tag{4.8}$$

with the weights estimate from equation 4.5.

4.3. ASSESSING BEAMFORMING QUALITY

Six measures are used to evaluate the performance of the beamformer. Speech intelligibility is evaluated by the objective intelligibility measures PESQ and STOI. The other measures that are used are the global and segmental SNR, the white noise gain and the array gain.

The PESQ and STOI algorithms both compare the original and processed signal. The chosen implementation of the PESQ is the ITU-T recommendation P.862 in 2002 [3]. A general overview of the PESQ, given by the ITU-T Recommendation P.862, is illustrated in figure 4.4. First, the delays between the original and processed signal are calculated to align the signals in time. Based on the delays, PESQ compares the signals using a perceptual model. This perceptual model is done by transforming the signals to an representation that is analogous to the psychophysical representation of audio signals in the human auditory system. The final step of the PESQ algorithm is using a cognitive model to compute two error parameters, one without and one with an asymmetry factor [25]. These parameters are combined to give an objective listening quality MOS. The output of the PESQ is a number in the range of -0.5 to 4.5, where -0.5 is the lowest quality and 4.5 is the highest quality.

The other intelligibility measure used to determine the performance of the beamformer was STOI. The used algorithm was proposed and implemented by taal et al. [24, 26].

The SNR is calculated by dividing the power of the signal by the power of the noise by the following equation:[34]

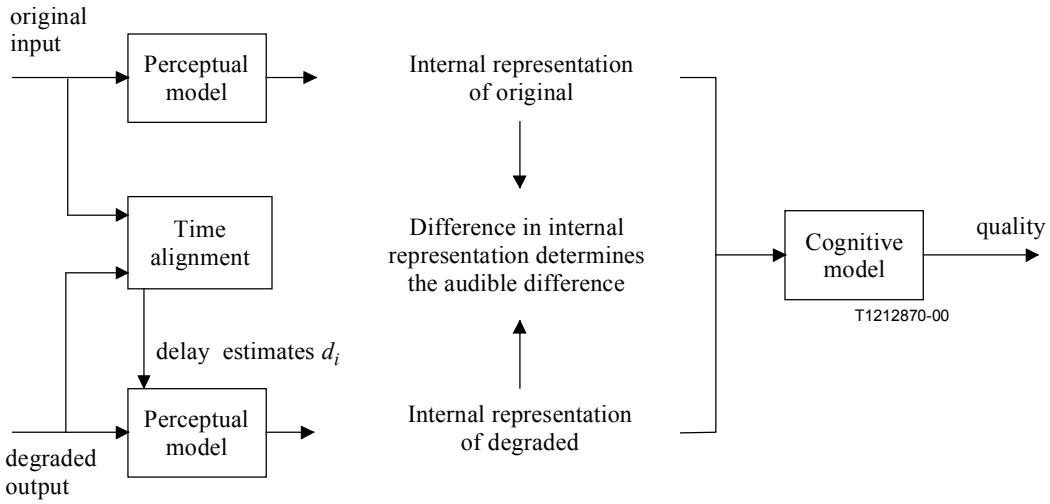
$$\text{SNR} = 10log_{10}\frac{\sum_{n=1}^{N} x^2(n)}{\sum_{n=1}^{N}\{x(n)-\hat{x}(n)\}} \tag{4.9}$$

Where x(n) is the clean speech signal, $\hat{x}(n)$ is the distorted speech and N is the number of samples.

The classical SNR is not well related to to the speech quality and therefore the segmental SNR (SNR_{seg}) will also be calculated. SNR_{seg} calculates the SNR in short frames and then takes the average of the frames.

$$SNR_{seg} = \frac{10}{K}\sum_{k=0}^{K-1} log_{10}\frac{\sum_{n=L_f k}^{L_f k+L_f-1} x^2(n)}{\sum_{n=L_f k}^{L_f k+L_f-1}\{x(n)-\hat{x}(n)\}^2} \tag{4.10}$$

Were L_f is the frame length and K the number of frames in the signal. By determining the energy in each frame it is possible to limit This calculation to the frames in which speech is active. The audio frames without speech will be discarded.



NOTE – A computer model of the subject, consisting of a perceptual and a cognitive model, is used to compare the output of the device under test with the imput, using alignment information as derived from the time signals in the time alignment module.

Figure 4.4: Overview of the PESQ algorithm [3]

4.4. REALTIME BEAMFORMING TOOLBOX

To make it easier to repeat the beamforming experiments a MATLAB toolbox has been developed. The approach of this toolbox is based on the Model-view-controller-model (MVC). Beamforming scripts (of which there are many available on the internet) can be coded into a standard processing unit to use them in real time. Changes to individual algorithms may have to be made to make them run in realtime and the input variables may have to be relabeled.
In this section the implementation of the system is described. Figure 4.5 shows a flowchart of the system. The green blocks (on the left side) are possible inputs. The orange blocks show inputs that have to be set from the UI. The rightmost blocks are the processing algorithms.

4.4.1. DATA ACQUISITION

Data inputs can be audio files, computer audio input channels, smartphones that run the Android app or multiple microphone locations can be simulated using different simulation options. Data from files can be loaded at once. The smartphones and computer microphones stream the data in small frames and call an update function when there is new data available. All inputs can be sampled at or resampled to other sample rates, but standard $48000Hz$ is used. The data is usually sampled with a resolution of 16 bits per sample, but as many steps in the computation require the data to be in double precision the data is stored in memory in double precision.

4.4.2. BUFFER

A fast data buffer has been implemented that serves as the core of the toolbox. All algorithms access the data so there need not be multiple copies of large data variables in memory. All of the used algorithms process the data in frames and write the processed data back to the buffer. This, in combination with proper memory allocation makes the toolbox run fast on large data sets as well as on smaller ones, as long as there is enough memory available. The buffer was tested with at most an hour of streaming audio from three smartphones, which resulted in about $3GB$ of memory use.
The properties of the data that are stored in the buffer are:

- Sample rate, $48000Hz$
- Bits per Sample, 16
- Number of Channels

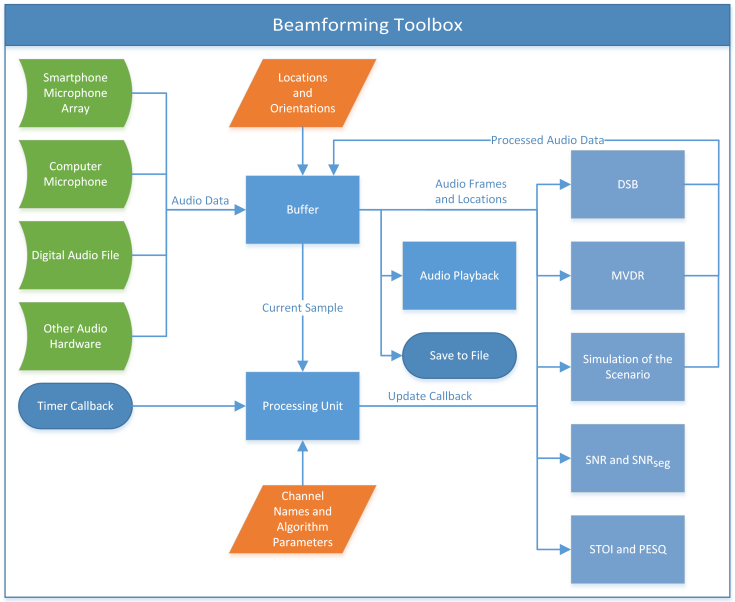


Figure 4.5: Beamforming Toolbox flowchart.

- Channel Name, reserved names are: $*Mic$,$*Source$,$*Noise$,$*Beamformer$
- Locations, $[x, y, z, az, el, up]$
- Current Sample, stores the index of the last sample in the buffer for each channel
- Total Samples
- Audio Data, matrix of size Number-of-Channels-by-Total-Samples

The locations are stored as vectors with in the first three entries the position in 3D, followed by the azimuth and elevation, and finally a boolean value for up or downward facing of the smartphone (screen). The latter three are only used for directional microphones and the last one is only used for smartphones.

4.4.3. SYNCHRONIZED RECORDING

Because the channels use multiple hardware devices the channels do not start the recording at the same time. A synchronized recording method has been developed to overcome this problem, similar to the method proposed by Bosma and Smeding [17]. When synchronized recording is started the computer adds a Maximum Length Sequence (MLS) at the beginning of the source recording. The MLS pulse has many nice properties but the most interesting one here is that the autocorrelation approaches a Kronecker delta [35]. Sufficient time for the synchronization to finish is also added before the source signal starts. The correlation between the MLS pulse in the source channel and the recorded MLS pulse is calculated and the highest value corresponds to the Kronecker delta, at the time offset between the signals. the time offset. This method is used on all the channels in the buffer to calculate the delay between when the first sample was taken and when the MLS pulse was received.
The TDOAs of all channels, calculated from locations inputted by the user, are substracted from this delay to force the TDOA of the channels to correspond to the locations.
A delayChannel method then shifts the samples in the buffer and the delays are added to the current sample value of the channel to ensure samples arriving in the future are lined up in time.

4.4.4. PROCESSING UNIT

Realtime processing of data which is arriving at irregular intervals and in irregular order has been made implicit by having the buffer store current sample values for each channel. The processing units check whether the current sample value of the listening channel plus the size of their frames is larger than the minimum of the current samples of each of the connected input channels. It will then loop over all available frames and put the processed data in their output channel in the buffer. This makes processing large files from the

disk go as fast as the computations will allow, and processing streaming data go in realtime for beamformers which don't have too large complexity. The MVDR beamformer complexity increases hugely with the number of input channels. For our purposes realtime means that the incoming and outgoing audiosignals, the beamformed signals and the SNR and SNR_{seg} can be plotted on screen with less than a second delay.

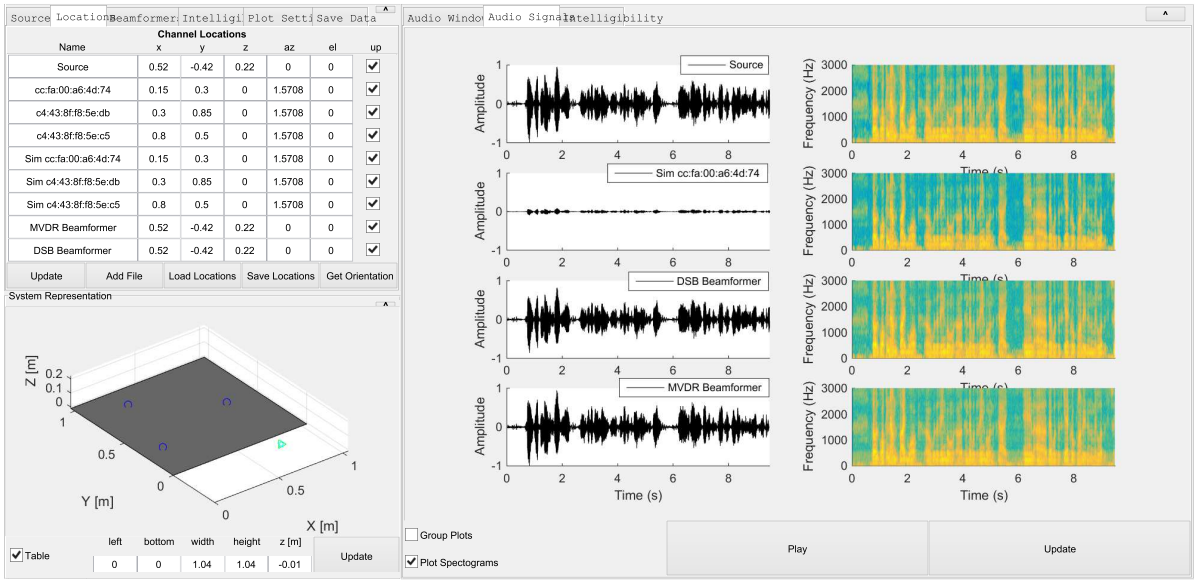


Figure 4.6: Matlab Realtime Beamforming Toolbox GUI

4.4.5. USER INTERFACE (UI)

The UI has been build with a modular design. Every UI panel can run individually, or they can be connected with a main object that has multiple UI panels. There is a panel where frames from the buffer can be plotted in realtime and one where the whole channel can be plotted. These can also be used to easily playback (part of) the data and to plot the frequency spectrum against time. There is a panel where the intelligibility metric outputs are shown. Each of the inputs has its own panel with buttons to start or stop streaming or several others. There is a panel which can be used to save data to multiple audio formats, with the locations stored in the meta-data of the audio file, or as MATLAB .mat files, with all relevant labels saved. Each of the beamformers has a panel to change their parameters and to enable them, or to process data at once. The intelligibility metrics have a similar panel and a separate panel where the metrics are shown, and the output of the SNR algorithm is plotted against time. There is a standard selector panel which is used in multiple places to connect inputs to processing units. All the locations can be entered in the locations panel. Lastly there is a system representation tab which plots the locations of the channels in the buffer on an axis and can also draw a tabletop surface. This makes it easy to compare modeled scenario with the real scenario.

5

RESULTS

There are two implemented MVDR algorithms, one assuming uniform microphone directivity and one including the directivities of the used microphones. Comparing these two gives an indication of what the influence of including the directivities are, since the directivities are the only difference between the two algorithms. It is expected that compensating for these directivities should lead to an increase in performance of the beamforming algorithm. This section covers the results of these experiments and simulations. These results are put in tables to give a clear overview of the results.

5.1. MEASUREMENT SCENARIO

A set of experiments have been done to demonstrate the effect of the different beamformers on recorded audio. Experiments excluding and including the directivities of the microphones have been done to determine the impact including directivities has on the beamforming algorithms. In this research, the DSB, the MVDR beamformer excluding directivities and the MVDR beamformer including directivities have been investigated.

There are two kinds of experiments conducted. The first are measurements were done in the anechoic chamber in the applied physics building at TU Delft. The other experiments are performed in an office room with dimensions 6.60m x 3.90m x 3.15m, $[x, y, z]$.

5.1.1. ANECHOIC CHAMBER EXPERIMENTS

Measurements have been done in the anechoic chamber to measure recorded audio without the presence of reverberations. Two Speakers and three smartphones were used during these measurements, as illustrated in figure 5.1. In this figure, the distances between the different objects are given.

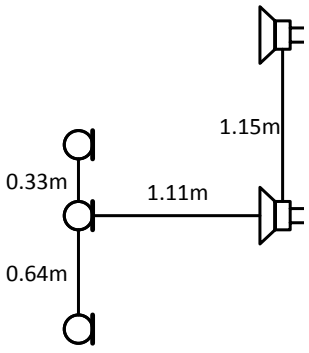


Figure 5.1: Anechoic setup

Beamforming has been performed on the recorded audio to determine the performance of the beamforming algorithms in a real room without reverberations. De audio signals and spectrograms of the source, the closest microphone and the different beamforming algorithms have been plotted in 5.2.

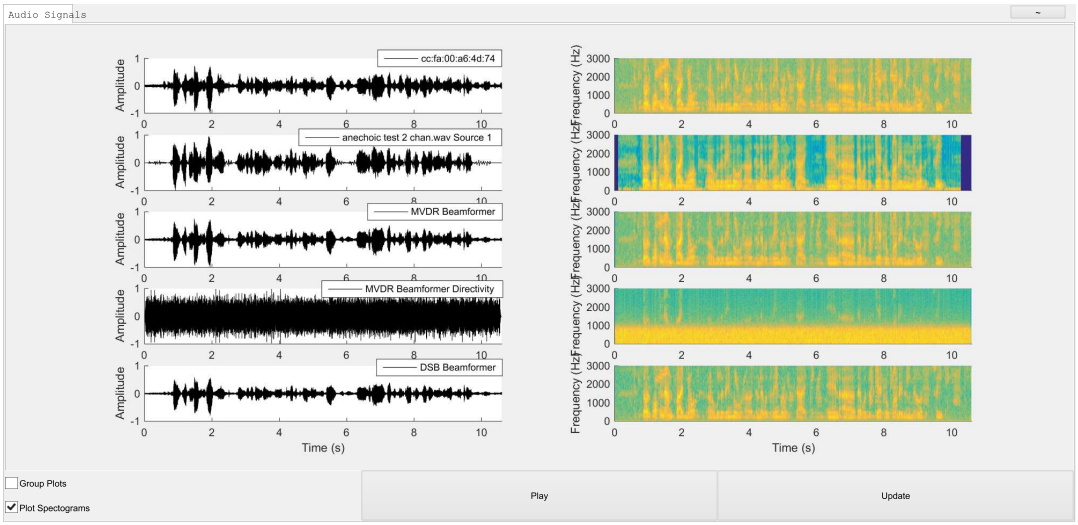


Figure 5.2: Audio signals and spectrogram anechoic chamber experiment 1: With an interfering audio source

These audio signals are used to calculate the intelligibility measures. The results of these intelligibility calculation are plotted in 5.3

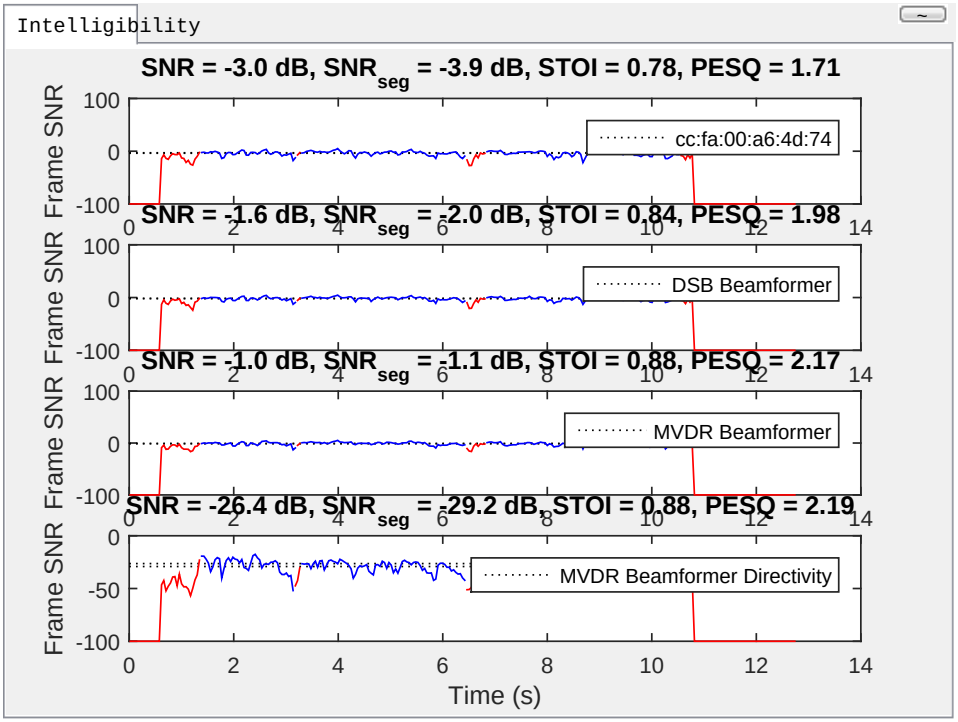


Figure 5.3: Intelligibility anechoic chamber experiment 1: With an interfering audio source

Finally the results have been put in table 5.1. The results have also been put in table 5.4, which is a large table at the end of the chapter with the anechoic chamber results as well as results from office room experiments. This table makes the different experimental results easily comparable.

	Anechoic measurement			
Beamformer	Closest	DSB	MVDR	MVDR+
Performance measure				
SNR	-3.0	-1.6	-1.0	-26.4
SNR_{seg}	-3.9	-2.0	-1.1	-29.2
STOI	0.78	0.84	0.88	0.88
PESQ	1.71	1.98	2.17	2.19
White noise gain		-4.65	56.85	102.05

Table 5.1: Experimental results of the anechoic chamber measurement

5.1.2. OFFICE ROOM EXPERIMENTS

During the office room experiments the sound is recorded with a sampling frequency of $48000\,Hz$ by smartphone microphones. This sound is used to perform the beamforming algorithms on.

During the office room experiments, the following procedure has been followed:

- 2 speakers were placed at known locations with English voices, one as interference and one as source.
- 3 smartphones placed at known locations with known orientations.
- Both speakers played audio of english voices at the same time.
- The smartphones streamed the audio data to the MATLAB toolbox where it was recorded and synchronized.

This configuration is illustrated in figure 5.4.

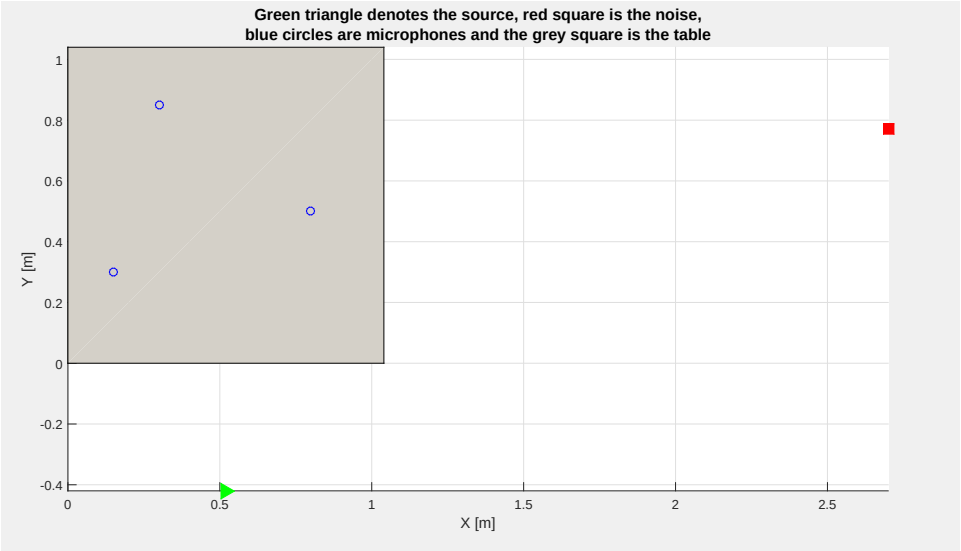


Figure 5.4: Office room experiment configuration. Microphone Locations: $(0.15, 0.3, 0)$, $(0.8, 0.5, 0)$, $(0.3, 0.85, 0)$, Source Location: $(0.52, -0.42, 0.22)$, Interferer Location: $(2.7, 0.77, -0.2)$

5.1.3. SIMULATIONS OF THE OFFICE ROOM EXPERIMENTS

Simulations of the same microphone setup as during the office room experiments have been done to see how much the conducted experiments differ from these simulated scenario. During the simulations, the parameters as described in section 4.1 have been varied to investigate the sensitivity of the beamforming algorithms to each of the parameters. The simulated scenarios are as follows:

Simulation	reverberations	Added white noise at SNR [dB]	Standard deviation of position errors
1	Excluded	60	0
2	Included	60	0
3	Excluded	40	0
4	Included	40	0
5	Excluded	60	0.05
6	Included	60	0.05

Table 5.2: The simulated scenarios

A perfect match cannot be expected since the simulated microphones are omnidirectional which is not the case for smartphone microphones. However, These simulations can determine if the beamforming algorithms perform in a simulated scenario and how sensitive the beamformers are to the different parameters.

For both the DSB and MVDR beamformer, the results of the simulations are put in table 5.3. For the fifth and sixth simulation, the SNR of the closest microphone has a negative value (in decibels). In this case, it is not representative for the performance of the beamformer to use the array-gain. This is because a positive SNR after a beamforming algorithm results in a negative array-gain and, in general, a negative array-gain indicates a poor performing beamforming algorithm. Therefore the entries in the table for the array gain are left blank for both these simulations.

5.1.4. RESULTS OF THE OFFICE ROOM EXPERIMENT

During the conducted office room experiments, the orientation of the microphones have been varied for the microphone setup earlier illustrated in fig 5.4. The investigated orientations are :

- $\theta = -90°$ (microphones facing the source)
- $\theta = 0°$ (microphones facing the interfering sound source)
- $\theta = 90°$ (microphones directed away from the source
- Each microphone in a different orientation.

Where $\theta = 0°$ is taken in the direction of the positive x-axis and the angle increases in the counterclockwise direction. In the remainder of this section, the microphone locations are denoted by their (x,y,z) location as illustrated in figure 5.4.

A scenario with different orientations was investigated since the aim of this thesis was to research beamforming algorithms with ad-hoc microphone arrays. The orientations of the microphones in an ad-hoc microphone array will also be different for each microphone. During this experiment, the microphone located at (0.3,0.85,0) was orientated at $\theta = -90°$, the microphone located at (0.8,0.5,0) was orientated at $\theta = 0°$ and the microphone located at (0.15,0.3,0) was oriented $\theta = 90°$.

The last office room experiment that has been conducted was with a linear array to compare the results of this array to the earlier ad-hoc arrays. The different microphones were located at different with the same y-coordinate. To be exact, the locations of the microphones were (0.27,0.75,0), (0.52,0.75,0), (0.77,0.75,0).

The office room experiments were once conducted with only the source speaker playing an English voice and once with the source speaker and noise speaker playing English voices. The different performance measures as presented in section 4.3 are calculated for each of the three beamforming algorithms and for the closest microphone. The performance measures of the closest microphone will be used to determine if and how much the signal is enhanced by the beamforming algorithms. With these results it is also possible to compare the result of each beamformimg algorithm and conclude which beamforming algorithm has the best performance. The results of the office room experiments are put in table 5.4. In this table, MVDR is the standard MVDR algorithm and MVDR+ stands for the MVDR beamforming algorithm with the compensation for the directional gain of the microphones included. The SNR of the closest microphones in each of the cases is negative (in decibels) and therefore the array-gain will not give a good indication about the performance of the beamforming algorithms. For this reason, the array-gain is left out of this table.

Table 5.3: Simulated results

	Simulation 1 Without reverberations White noise added at 60dB SNR Without position errors			Simulation 2 With reverberations White noise added at 60dB SNR Without position errors			Simulation 3 Without reverberations White noise added at 40dB SNR Without position errors		
Beamformer	Closest	DSB	MVDR	Closest	DSB	MVDR	Closest	DSB	MVDR
Performance measure									
SNR	6.2	12.6	34.7	2.9	8.2	8.6	5.2	11.6	17.1
SNR_{seg}	4.8	10.1	28.7	-1.3	3.5	5.6	1.5	7.5	11.4
STOI	0.80	0.90	0.99	0.77	0.88	0.95	0.78	0.89	0.93
PESQ	1.82	2.29	3.66	1.83	2.16	2.92	1.65	2.09	2.28
Array gain		2.03	5.61		2.78	2.93		2.22	3.27
White noise gain		-4.64	51.92		-4.64	52.69		-4.64	50.94

	Simulation 4 With reverberations White noise added at 40dB SNR Without position errors			Simulation 5 Without reverberations White noise added at 60dB SNR with position errors with a sigma of 0.05			Simulation 6 With reverberations White noise added at 60dB SNR with position errors with a sigma of 0.05		
Beamformer	Closest	DSB	MVDR	Closest	DSB	MVDR	Closest	DSB	MVDR
Performance measure									
SNR	2.5	7.8	9.5	-2.4	1.6	-2.3	-3.3	-0.1	-0.6
SNR_{seg}	-2.1	3.0	4.9	-6.3	-3.0	-6.5	-7.3	-4.5	-3.3
STOI	0.75	0.87	0.91	0.68	0.79	0.85	0.70	0.76	0.86
PESQ	1.66	2.00	2.25	1.55	1.91	2.41	1.75	1.71	2.50
Array gain		3.19	3.88		-0.64	0.94			
White noise gain		-4.64	50.96		-4.63	50.93		-4.64	51.05

Table 5.4: Experimental results

Without an interfering audio source												
	$\theta = -90°$				$\theta = 0°$				$\theta = -90°$			
Beamformer	Closest	DSB	MVDR	MVDR+	Closest	DSB	MVDR	MVDR+	Closest	DSB	MVDR	MVDR+
Performance measure												
SNR	-6.5	-5.9	-9.4	-32.1	-6.7	-6.2	-14.3	-47.3	-6.5	-5.8	-9.5	-33.3
SNR_{seg}	-3.0	-1.8	-6.4	-31.7	-7.0	-6.0	-15.5	-49.9	-6.6	-5.4	-10.7	-35.9
STOI	0.73	0.81	0.73	0.74	0.71	0.76	0.50	0.51	0.71	0.76	0.58	0.57
PESQ	2.45	2.68	2.18	2.24	2.34	2.53	1.47	1.55	2.36	2.54	1.78	1.74
White noise gain		-4.64	50.99	105.06		-4.63	52.24	96.59		-4.64	52.21	93.80
	Different orientations				Linear array							
Beamformer	Closest	DSB	MVDR	MVDR+	Closest	DSB	MVDR	MVDR+				
Performance measure												
SNR	-6.4	-5.8	-9.8	-34.7	-7.1	-8.6	-13.0	-55.3				
SNR_{seg}	-6.3	-5.2	-10.8	-37.3	-7.0	-11.0	-14.1	-57.9				
STOI	0.70	0.78	0.58	0.58	-13.0	0.51	0.47	0.48				
PESQ	2.36	2.59	1.71	1.74	-55.3	2.04	1.53	1.70				
White noise gain		-4.64	52.66	96.19		-4.77	54.08	105.10				

With an interfering audio source												
	$\theta = -90°$				$\theta = 0°$				$\theta = 90°$			
Beamformer	Closest	DSB	MVDR	MVDR+	Closest	DSB	MVDR	MVDR+	Closest	DSB	MVDR	MVDR+
Performance measure												
SNR	-6.8	-5.8	-7.0	-29.9	-7.3	-6.3	-8.9	-31.2	-6.7	-6.5	-13.8	-53
SNR_{seg}	-6.7	-5.0	-7.8	-32.5	-7.8	-6.1	-9.7	-33.8	-7.3	-7.6	-15.7	-55.6
STOI	0.70	0.79	0.64	0.65	0.66	0.72	0.56	0.58	0.65	0.65	0.48	0.47
PESQ	2.32	2.54	1.81	1.89	2.10	2.37	1.72	1.83	2.06	1.97	1.29	1.39
White noise gain		-4.64	52.82	105.08		-4.64	52.33	96.53		-4.64	51.96	93.82
	Different orientations				Linear array				Anechoic measurement			
Beamformer	Closest	DSB	MVDR	MVDR+	Closest	DSB	MVDR	MVDR+	Closest	DSB	MVDR	MVDR+
Performance measure												
SNR	-6.8	-6.0	-7.9	-28.5	-7.1	-6.2	-7.2	-34.1	-3.0	-1.6	-1.0	-26.4
SNR_{seg}	-7.5	-5.9	-8.8	-31.1	-7.6	-6.2	-8.1	-36.7	-3.9	-2.0	-1.1	-29.2
STOI	0.64	0.73	0.58	0.60	0.67	0.70	0.53	0.54	0.78	0.84	0.88	0.88
PESQ	2.08	2.36	1.72	1.81	2.12	2.30	1.68	1.80	1.71	1.98	2.17	2.19
White noise gain		-4.64	52.24	98.88		-4.77	54.56	104.99		-4.65	56.85	102.05

6

DISCUSSION

Different experiments and simulations have been done to investigate the performance of the beamforming algorithms. The DSB is a simple beamforming algorithm compared to the MVDR beamforming algorithm and as discussed in section 3, the DSB algorithm is optimal for spatially white noise. The delay-and-sum beamforming algorithm can be used as comparison for the other two beamforming algorithms. The MVDR algorithm was expected to perform well in the presence of coherent noise. Therefore it was expected that the MVDR algorithm would be the best performing beamformer in an office room scenario.

6.1. PERFORMANCE OF THE BEAMFORMING ALGORITHMS

The simulated scenarios confirm the expectation that the MVDR beamforming algorithm has the best performance, but the conducted experiments lead to unexpected results. The DSB algorithm slightly enhanced the quality of the audio in the simulated scenarios and in the office room experiments but the MVDR only improved the sound quality in the simulated scenarios and in the anechoic scenario. A possible explanation for this could be the high white noise gain of the MVDR beamforming algorithm. In all the simulations and all the conducted measurements, the white noise gain for the MVDR beamformer was above $50dB$. A high white noise gain causes the beamforming algorithm to have a poor SNR due to white noise contributions as explained in chapter 3. The MVDR has a tendency to converge to the DSB when the white noise level is high [32]. Another reason could be the sensitivity of the MVDR to position errors. As proven in the simulations, the MVDR is highly sensitive to position errors. Inaccurately measured locations during the office room experiments could reduce the performance of the MVDR beamformer. Also inacuracy in speed of sound could be a factor in the performance decrease. The speed of sound varies with temperature and humidity, and this variation was not considered in the measurements.

6.2. POSSIBLE IMPROVEMENTS

Localization and distributed calculation of the beamforming algorithms are not in the scope of this thesis, but are still closely related topics. The source and microphone locations are essential for the performance of a beamforming algorithm. In this research, there is still a central computer needed to run the beamforming algorithm. Hence it would be an improvement if the calculations of the beamforming algorithm could be done distributed on the smartphones. These distributed calculations could make a central processing unit redundant.

6.2.1. LOCALIZATION

Brandstein et al. [8] divide source localization in three different classes, steered beamformer based locators, high resolution spectral estimation locators and time difference of arrival (TDOA) locators. Steered beamformer based locators scan the region to find the region which produces the highest output energy. High-resolution spectral estimation locators use a spectral phase correlation matrix to find the best fit for source localization. TDOA locators use the delay of the sound between microphones to calculate the location of the source.

The locations of the microphones are, besides the location of the source, also of importance to the performance of the beamformer. A possible solution for this problem is the use of self-localization with high

accuracy to find out the locations of the microphones [36].

6.2.2. DISTRIBUTED ALGORITHM

A distributed algorithm for the DSB for wireless acoustic sensor networks, is proposed by Zeng et al. [37, 38]. A distributed MVDR beamformer for wireless acoustic sensor networks based on message passing is proposed in [39].

6.2.3. REALTIME BEAMFORMING TOOLBOX

The MATLAB toolbox developed for this project was very powerful in conducting the experiments and analyzing the data. The complete process has been automated. Future work on audio enhancement using smartphones could use this toolbox to record audio synchronously on smartphones.
Also the opportunity to add more different beamforming algorithms can be explored. There is an example processing class in which they can be included quite easily.

7

CONCLUSION

The goal of the research was to compare the performance of existing wide-band near-field beamforming algorithms in enhancing audio from microphone arrays, and the adaptation of directivity patterns in the microphone response. Measurements have been performed both in an anechoic environment and in a typical office environment. Data has been gathered and processed using the beamforming toolbox written for this project. The data was analyzed and compared in both objective and subjective metrics, as is common in the literature about this subject [1, 8, 11–15, 40]. The objective audio quality measures show very high correlation with each other. They also show high correlation to the power ralated metrics, except when they are used on a very noisy signal. In that case both STOI and PESQ become unreliable.

7.1. SIMULATION

The simulations show improvement on all metrics for both DSB and MVDR in the measurement scenario. They also show better intelligibility for MVDR than for DSB in all scenarios, for both STOI and PESQ.
When the scene is simulated with low noise, no reverberations and no error in the locations the MVDR shows huge improvement in the SNR_{seg} of up to $24dB$. This is expected because the MVDR beamforming algorithm is statistically optimal. The DSB shows a more moderate improvement of about $5dB$ in the same scenario. When the reverberations are also simulated SNR_{seg} for DSB stays about equal. This is expected because the DSB is good at canceling spatially white noise. The SNR_{seg} of the MVDR will drop down to about the same level as the DSB when reverberations are simulated. When errors in the localization are simulated the performance increase almost vanishes. These results have the same trend as results by Martínez et al. [15] This means that both beamformers are not robust in localization errors.

7.2. MEASUREMENTS

Measurements on a typical setup for teleconferencing with one English speaking source and an English speaking interferer and office background noise show that in our particular setup the MVDR never beats the performance of the DSB for all the presented scenarios and metrics.

In general, the results of the DSB show a moderate improvement in both the power related measures and the intelligibility measures. The DSB showed a consistant performance these measurements. In contrary to the DSB, the MVDR did not improve either of the metrics in the measured scenarios.

The white noise gains for each beamformer is very constant in for all scenarios in both simulations and measurements. The average white noise gain for the DSB is $-4.6dB$, for the MVDR $52dB$ and MVDR with Directivity $98dB$. These values indicate that the MVDR algorithm will amplify spatially white noise much more than desired.

7.3. LOCATION ERRORS

It is a well known fact in literature that the MVDR beamformer performance degrades significantly when there are errors in the estimated locations. The simulations are able to reproduce this effect. The DSB beamformer also shows degradation of the Intelligibility and SNR when there is a simulated normal distributed position

error with a standard deviation of $5cm$ on every coordinate, the MVDR shows irregular results in measurement scenarios. Possibly, this is the reason for the difference in performance between the simulation and the measurement scenarios.

7.4. DIRECTIVITY

Because the MVDR without directivity does not show improvement in our measurement scenario, it may be questionable how useful the results about the MVDR with directivity are. And as the simulations do not use microphone directivity responses the performance can not be compared with the simulations. However, we can see that including the directivities as described in section 4.2.3 delivers mostly noise. There can be multiple reasons for this. It may be that there is an error in the implementation of the algorithm or it may be that the directivity impulse responses are not used in the correct way or the smartphones might have too erratic directivity patterns. This setup unfortunately does not show the same promising results as the literature does on the directivity measurements.

A

ETHICAL CONSIDERATIONS

This chapter in the thesis is not technically part of the thesis. It describes some ethical considerations related to (smartphone) microphone arrays and their applications. This chapter is included in all thesis works of the project group, e.g. also the work of Brinkman and de Rooij [16] and Bosma and Smeding [17]. This chapter contains *opinions* of the authors and is not a technical document.

A.1. THE NEED FOR ETHICAL DISCUSSION

In creating the beamforming system, we envisioned a single purpose: speech enhancement in ad-hoc teleconference calls. Unfortunately, history teaches us some well-intended inventions turn out to be dangerous to the point where regulation is needed to ban it. Examples are numerous: tetraethyllead was added to gasoline starting in the 1920s but phased out worldwide from the 1970s because of its environmental impact. Radithor is another 1920s example; a "wonder medicine" in its heyday, it was subsequently discovered to be extremely toxic to the point that the main marketeer for the drugs' "jaw fell off". Although these examples are very extreme, it is worthwhile examining the potential influence on society of our system-to-be.

A.2. UNINTENTIONAL USE CASES

The question we must ask ourselves thus becomes: what are some possible unforeseen use cases of our product? The possibility of a surveillance state (or company) using ad-hoc beamforming techniques to snoop on all its residents springs to mind. The revelations of NSA whistle blower Edward Snowden have documented just how far some American and European states are willing to go to keep tabs on their citizens. Moreover, they have revealed great levels of cooperation from tech-industrial giants such as Google, Facebook, Apple and others.

A.3. BEAMFORMER CONTRIBUTION

But how much would a smartphone beamforming system contribute to any level of mass-surveillance? The truth is that beamformers only work when the locations of smartphones are known to extreme accuracy. Such accuracy is not achievable with sensors fitted on even the latest smartphones. Furthermore, indoor localization of smartphones has thus far always relied on either large, extra hardware (RF beacons) or a sonic beacon. Neither seem appropriate to a state wanting to covertly listen in on conversations.

A.4. APPLICATIONS

Our prime focus has been with companies deploying the technology in teleconferencing situations. We do not seek out to be the next Skype, Hangouts or appear.in. A company using such software must be fully aware of the consequences of privacy invasion or potential information leakage caused by such software. Our solution would not be connected directly to the internet; it could, for example, serve as a virtual microphone input to the computer. What is done with the recorded audio is completely up to the user.

A.5. FINAL REMARKS

Theoretically, if the technology becomes more mature, it may become a weapon in the arsenal of surveillance states. But the technology is not remotely there yet. Ad-hoc beamforming on smartphones is still relatively uncharted territory and the lack of silent, unnoticable synchronization options kills off a potential to listen in discretely. We honestly see no potential for abuse of the technology investigated in our research.

B

DETAILED RESULTS

B.1. SIMULATION

Simulation	reverberations	Added white noise at SNR [dB]	Standard deviation of position errors
1	Excluded	60	0
2	Included	60	0
3	Excluded	40	0
4	Included	40	0
5	Excluded	60	0.05
6	Included	60	0.05

Table B.1: The simulated scenarios

B.1.1. SIMULATION 1

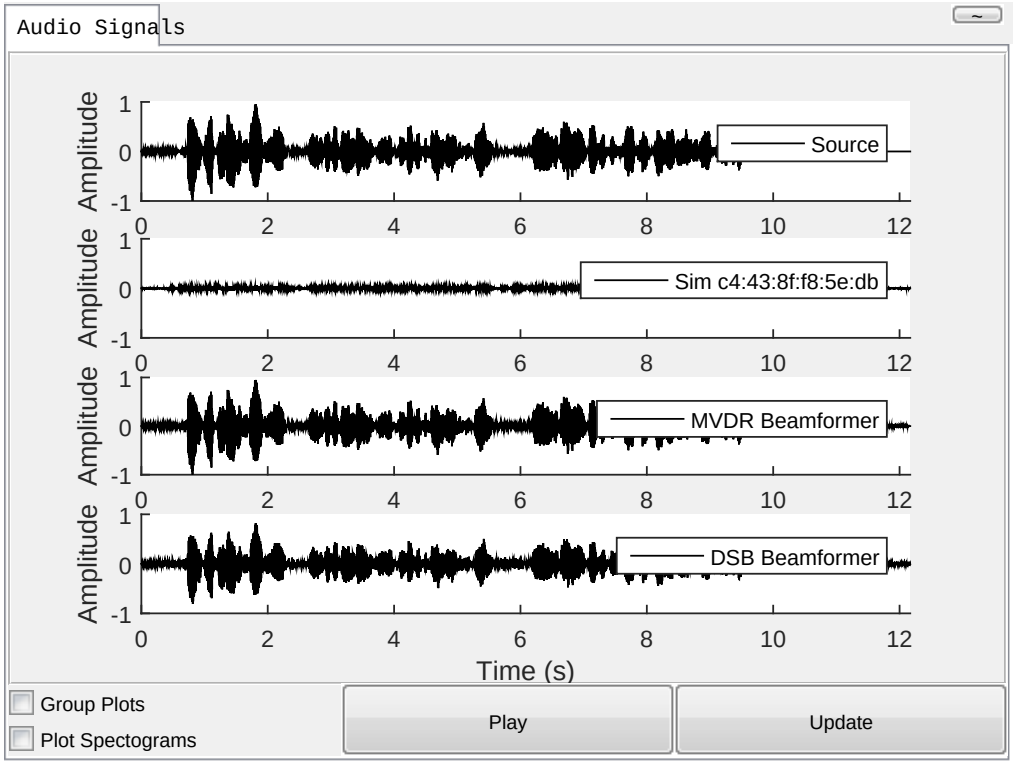


Figure B.1: Audio signals simulation 1: Without reverberations, White noise added at 60dB SNR, without position errors

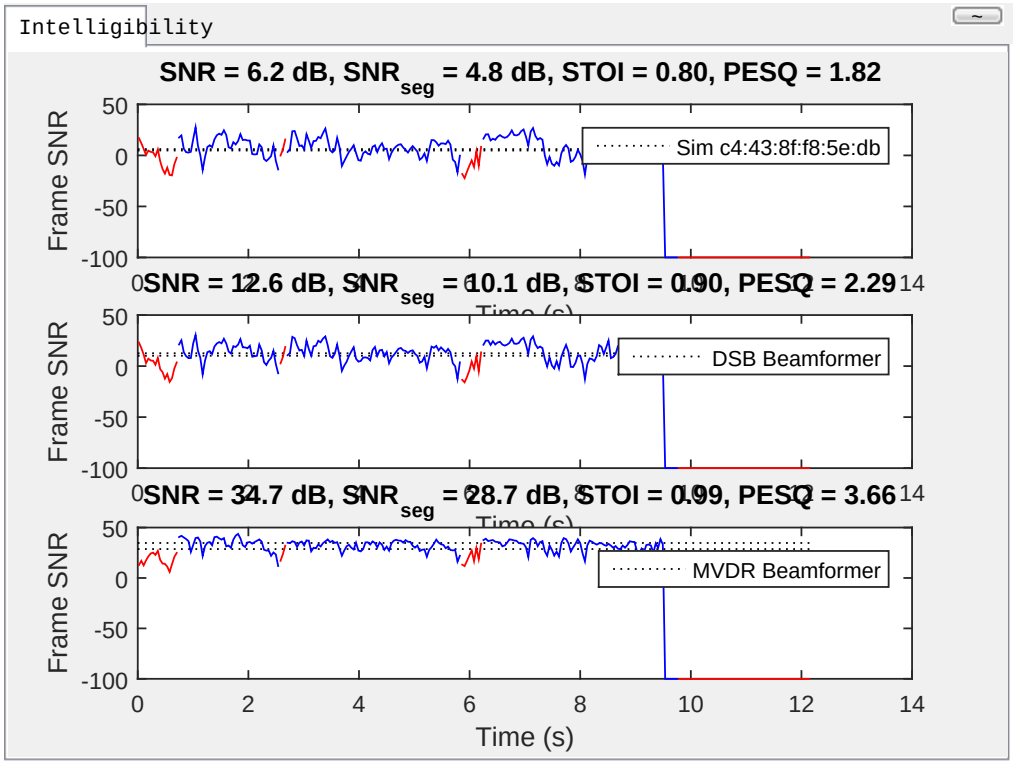


Figure B.2: Intelligibility simulation 1: Without reverberations, White noise added at 60dB SNR, without position errors

B.1.2. SIMULATION 2

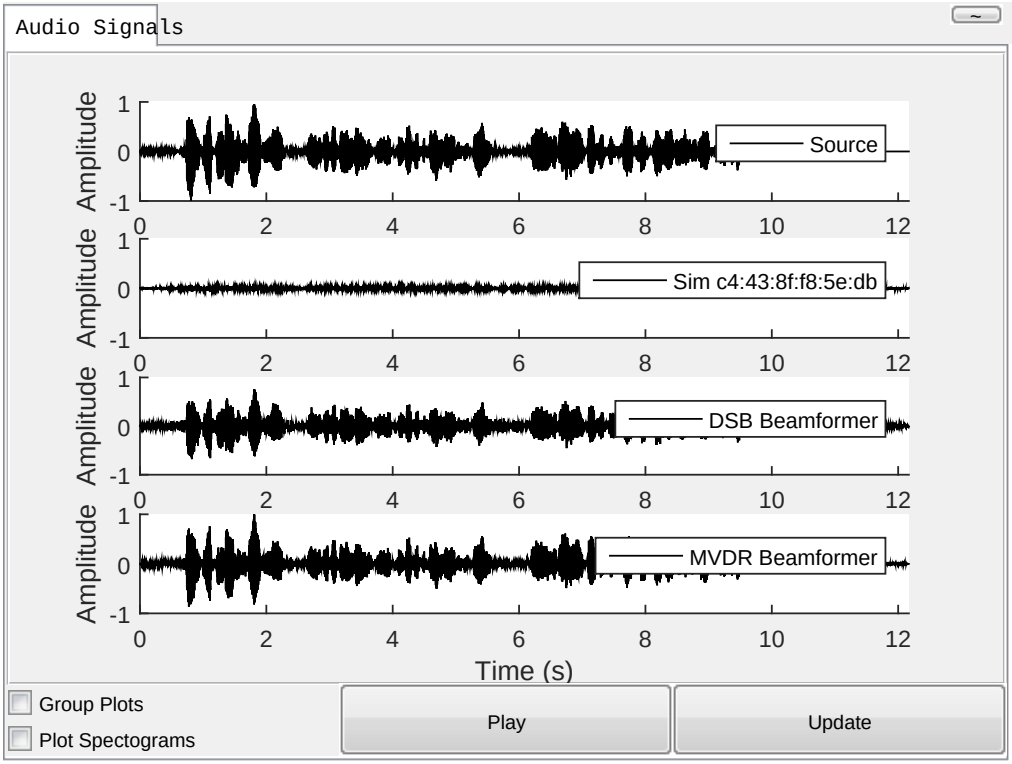


Figure B.3: Audio signals simulation 2: With reverberations, White noise added at 60dB SNR, without position errors

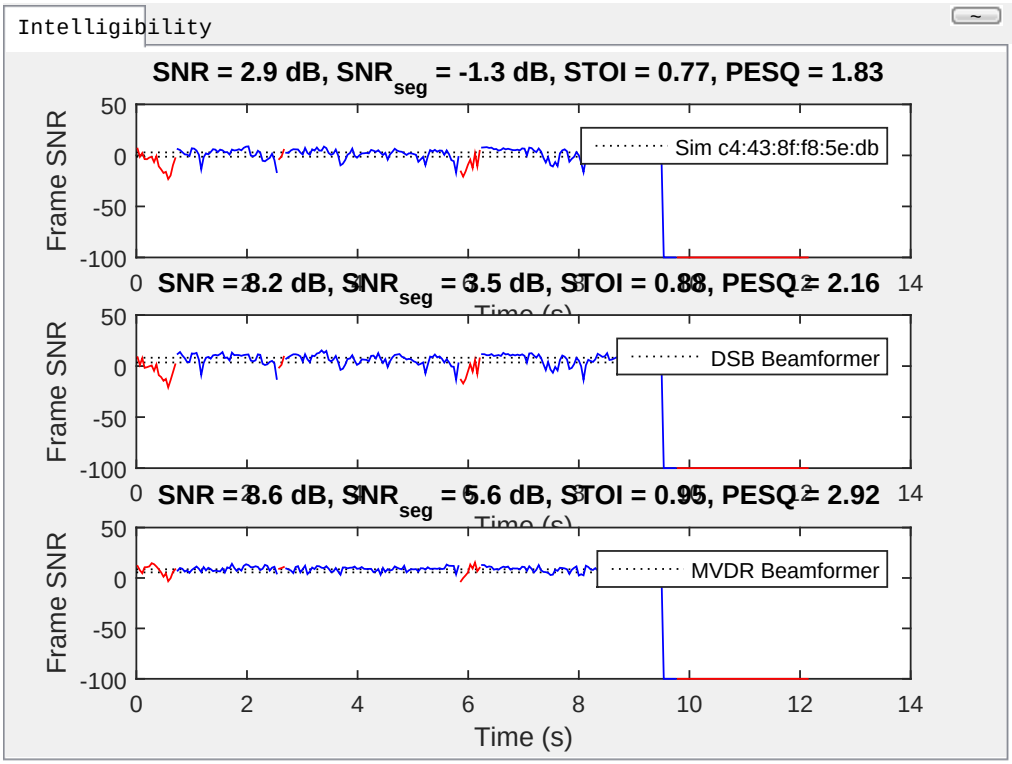


Figure B.4: Intelligibility simulation 2: With reverberations, White noise added at 60dB SNR, without position errors

B.1.3. SIMULATION 3

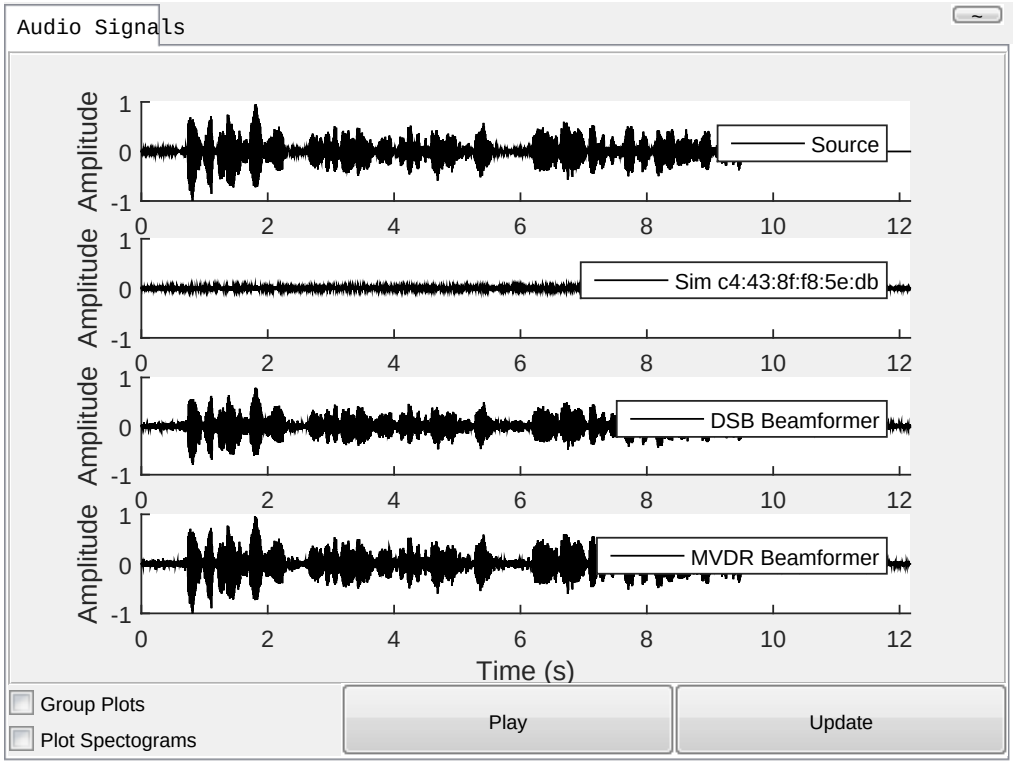


Figure B.5: Audio signals simulation 3: Without reverberations, White noise added at 40dB SNR, without position errors

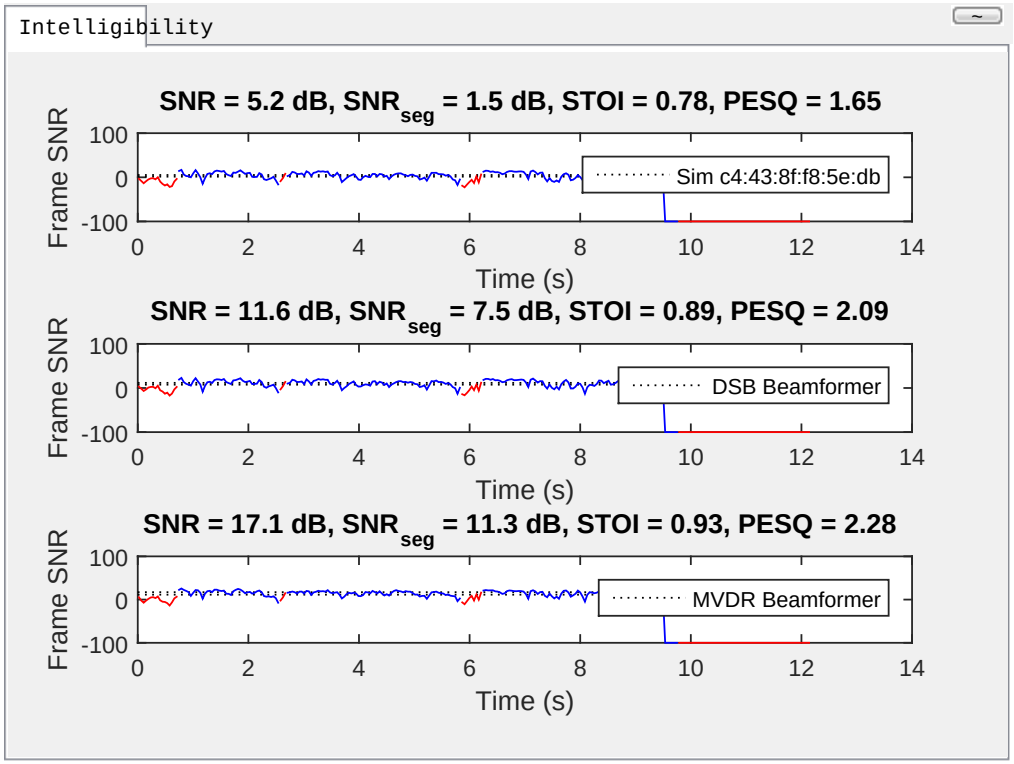


Figure B.6: Intelligibility simulation 3: Without reverberations, White noise added at 40dB SNR, without position errors

B.1.4. SIMULATION 4

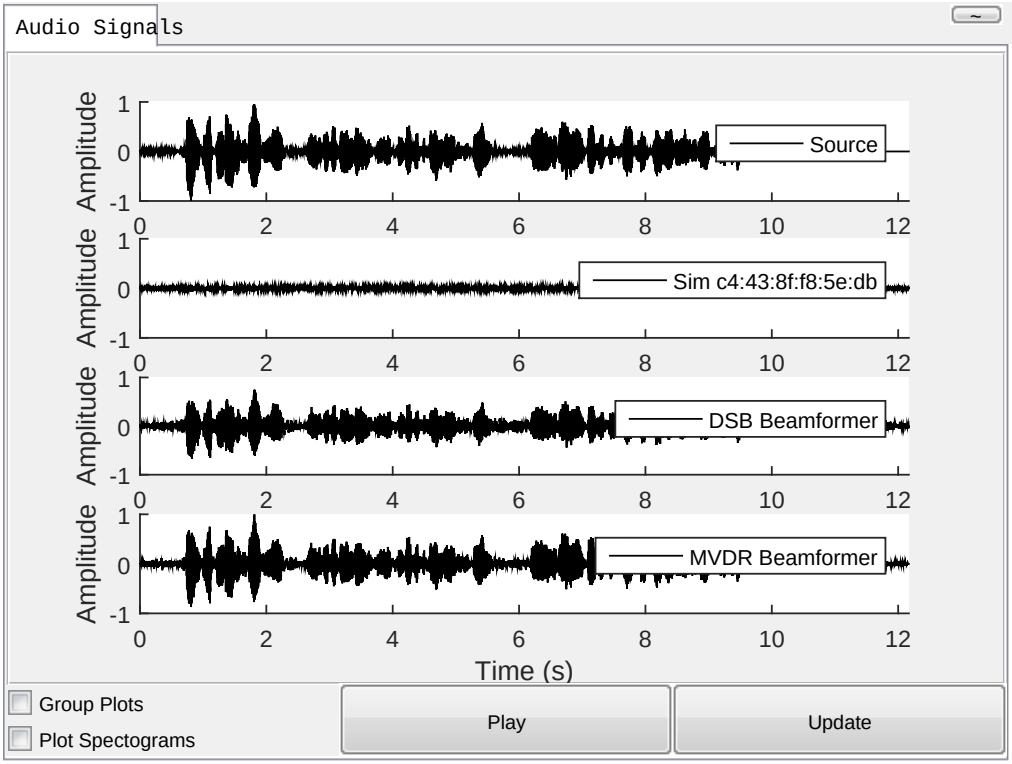


Figure B.7: Audio signals simulation 4: Without reverberations, White noise added at 40dB SNR, without position errors

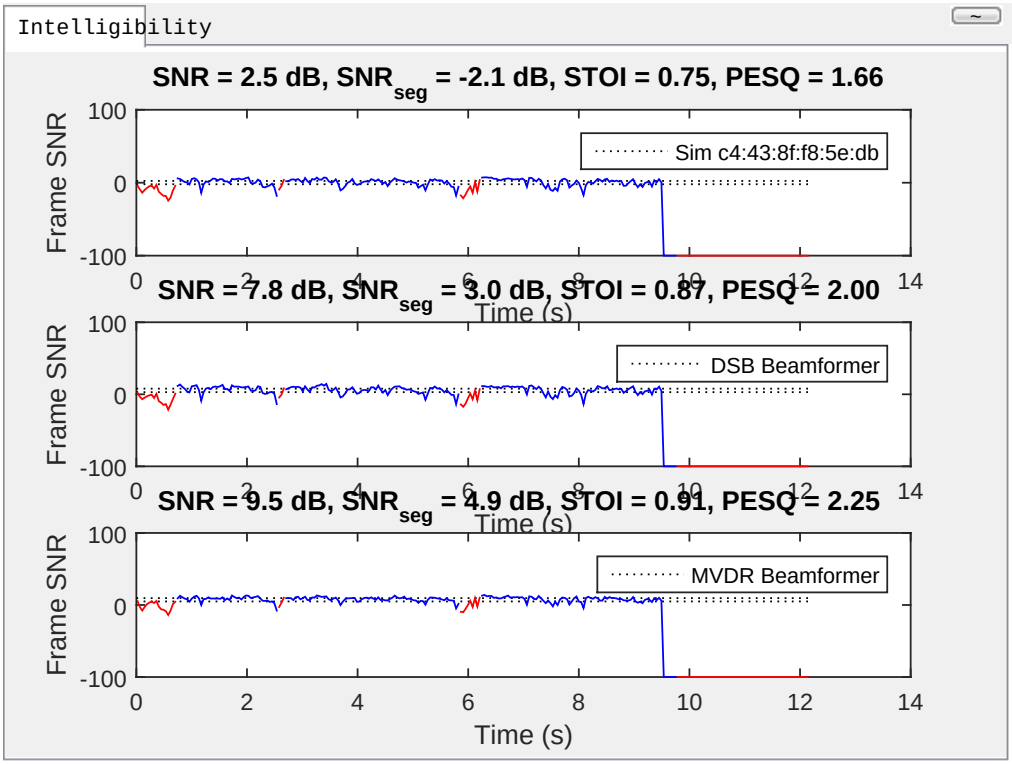


Figure B.8: Intelligibility simulation 4: Without reverberations, White noise added at 40dB SNR, without position errors

B.1.5. SIMULATION 6

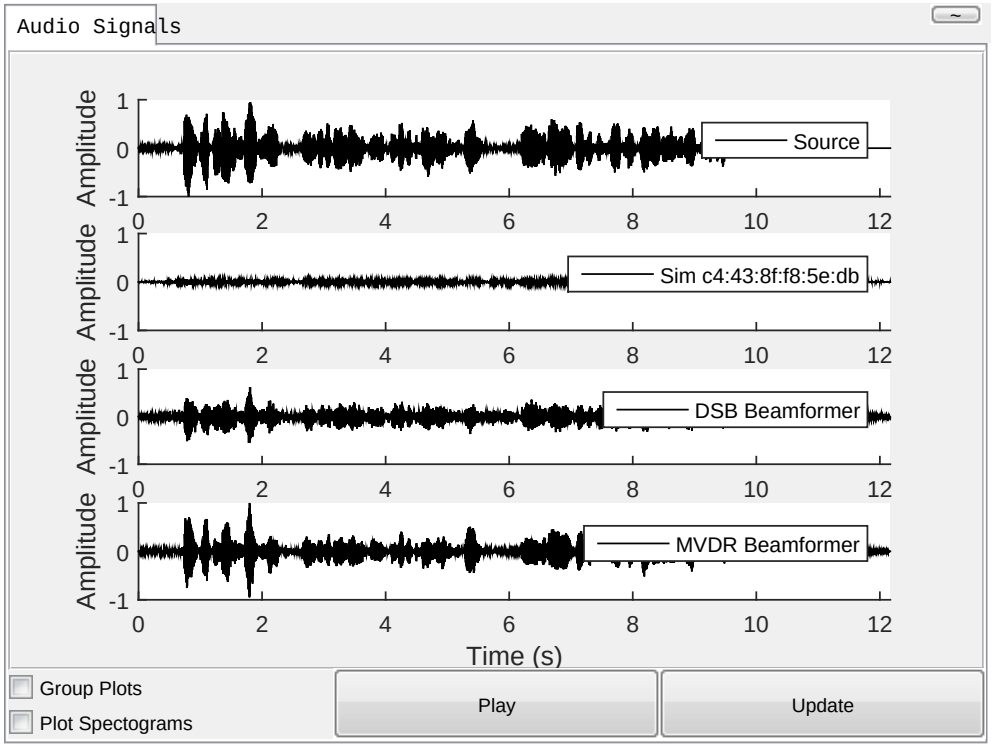


Figure B.9: Audio signals simulation 6: With reverberations, White noise added at 60dB SNR, with position errors with a sigma of 0.05cm

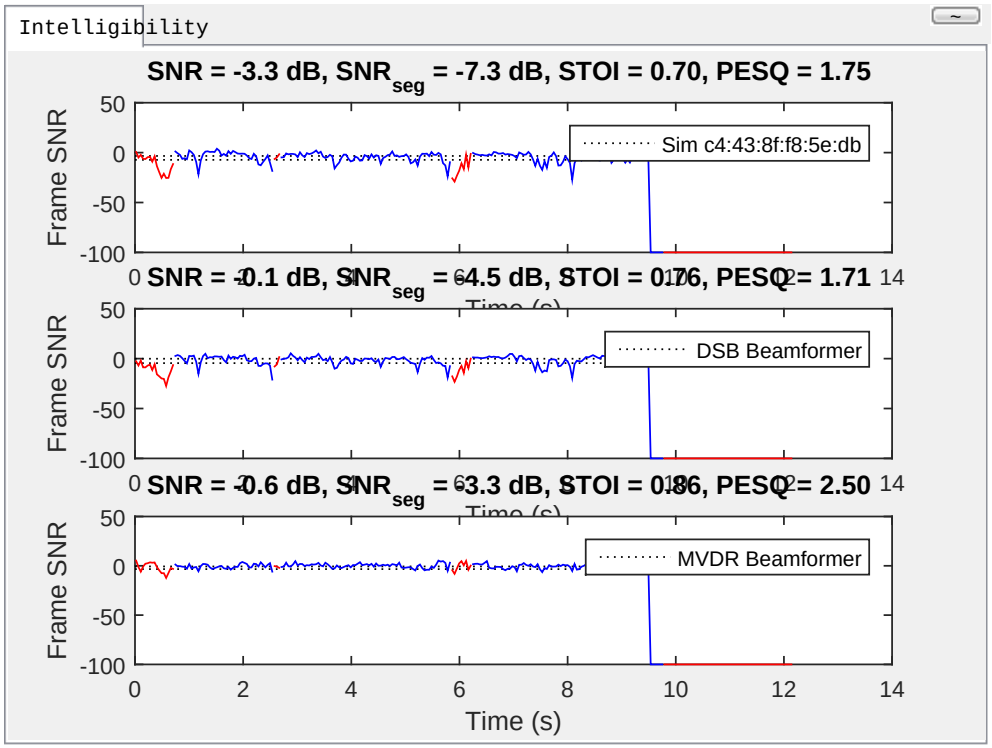


Figure B.10: Intelligibility simulation 6: With reverberations, White noise added at 60dB SNR, with position errors with a sigma of 0.05cm

B.2. OFFICE ROOM EXPERIMENTS

B.2.1. OFFICE ROOM EXPERIMENT 1

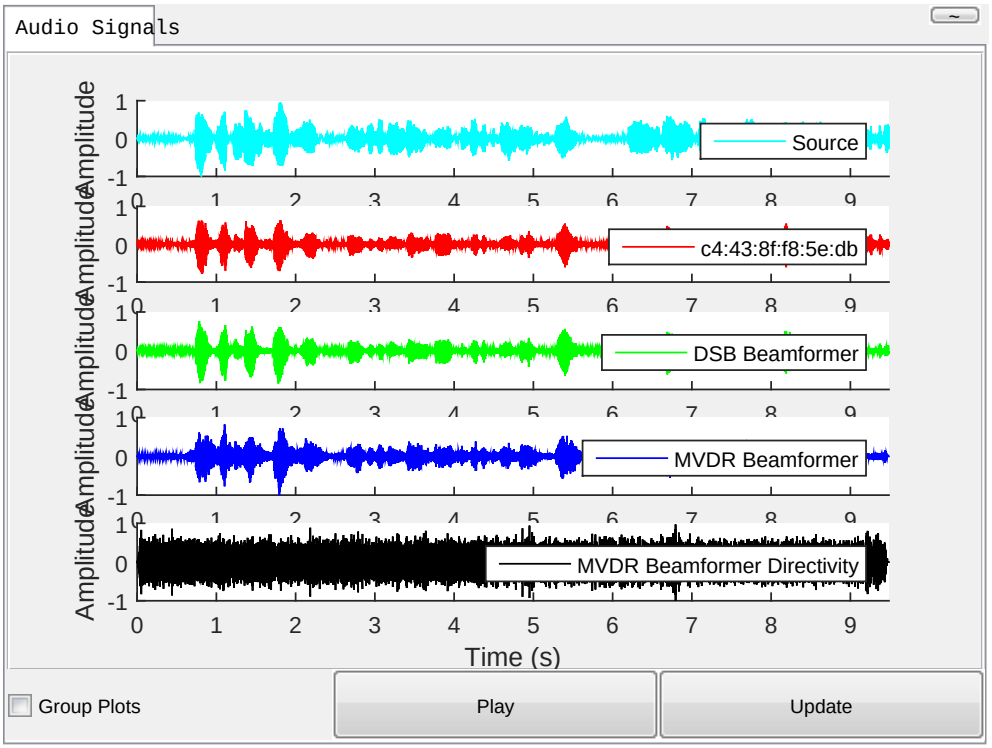


Figure B.11: Audio signals office room experiment 1: With an interfering audio source, $\theta = -90°$

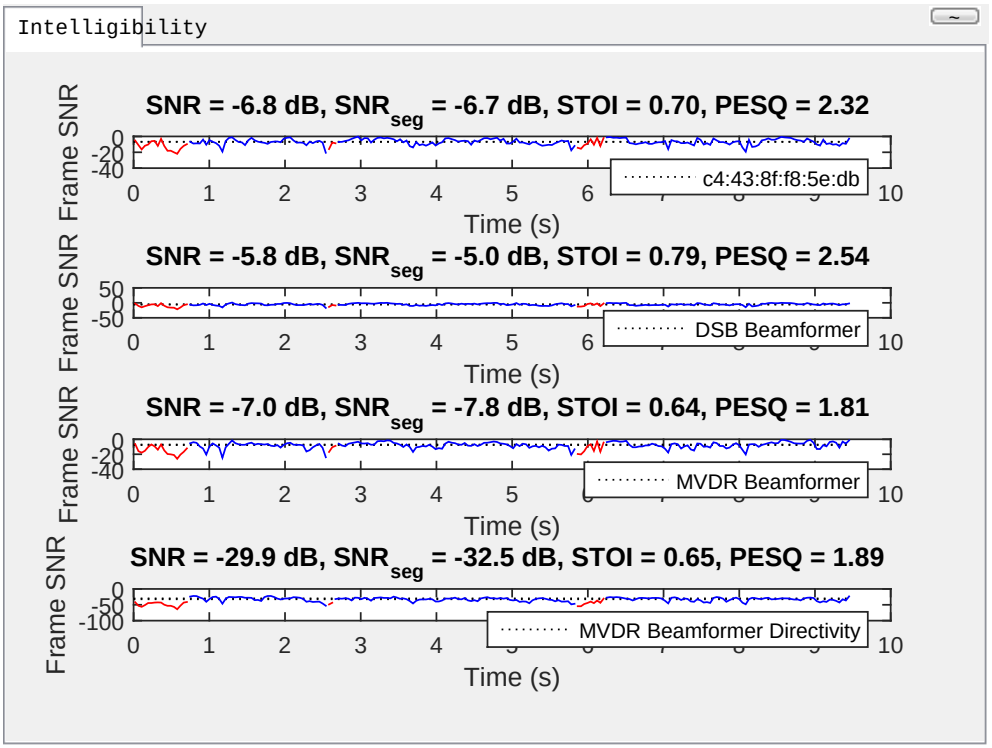


Figure B.12: Intelligibility office room experiment 1: With an interfering audio source, $\theta = -90°$

B.2.2. OFFICE ROOM EXPERIMENT 2

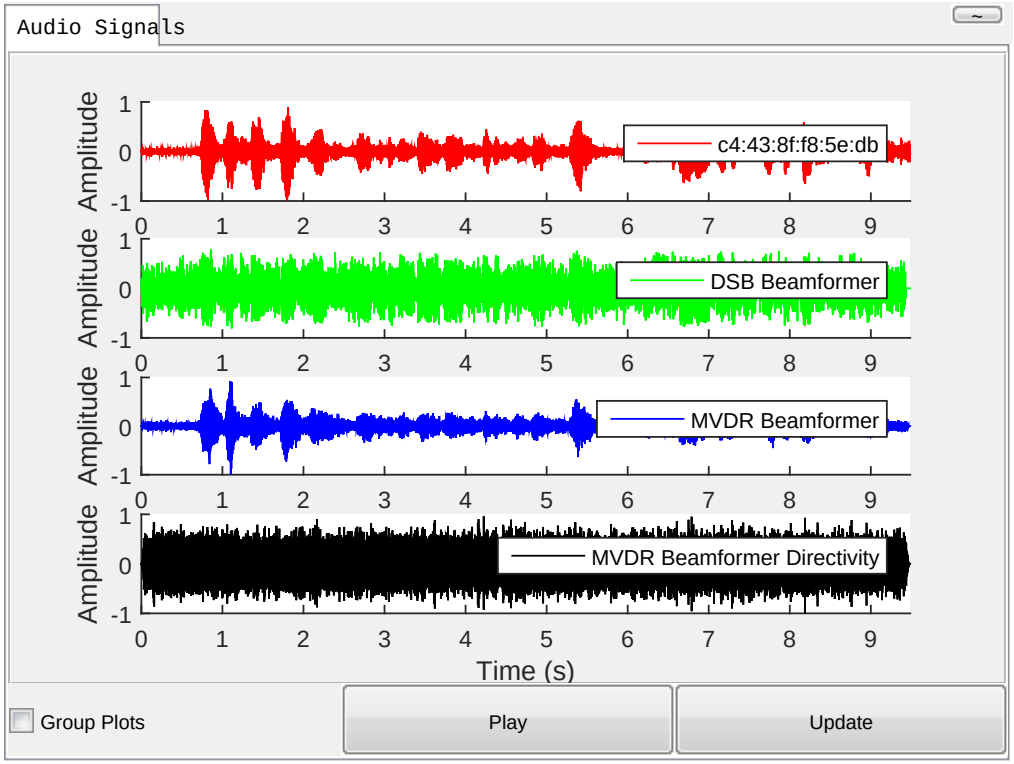


Figure B.13: Audio signals office room experiment 2: With an interfering audio source, $\theta = 0°$

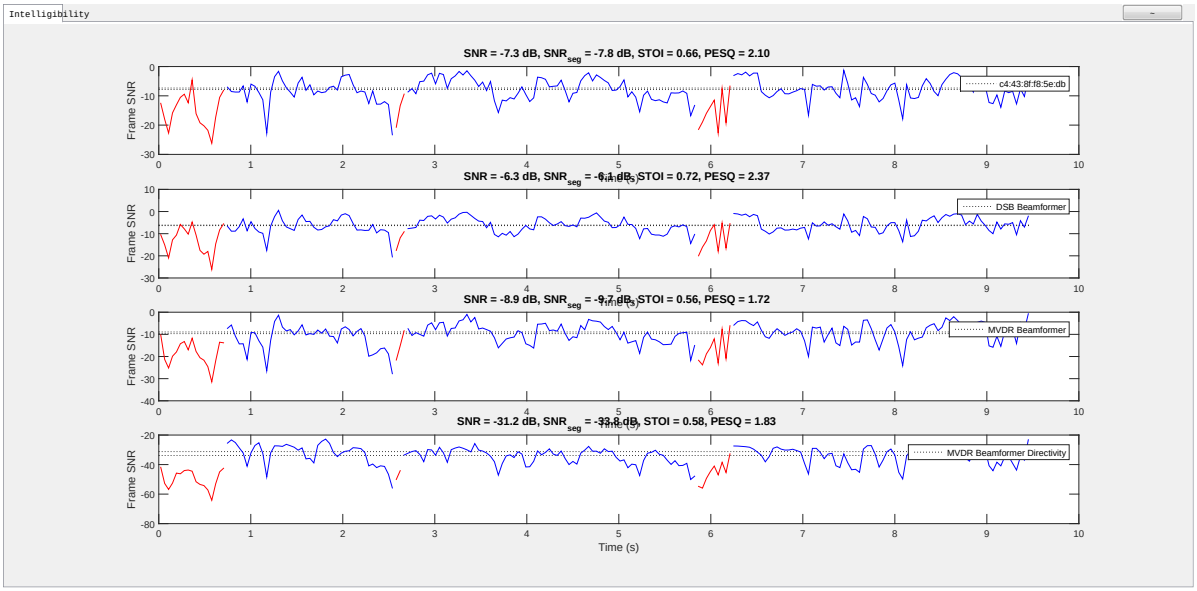


Figure B.14: Intelligibility office room experiment 2: With an interfering audio source, $\theta = 0°$

B.2.3. OFFICE ROOM EXPERIMENT 3

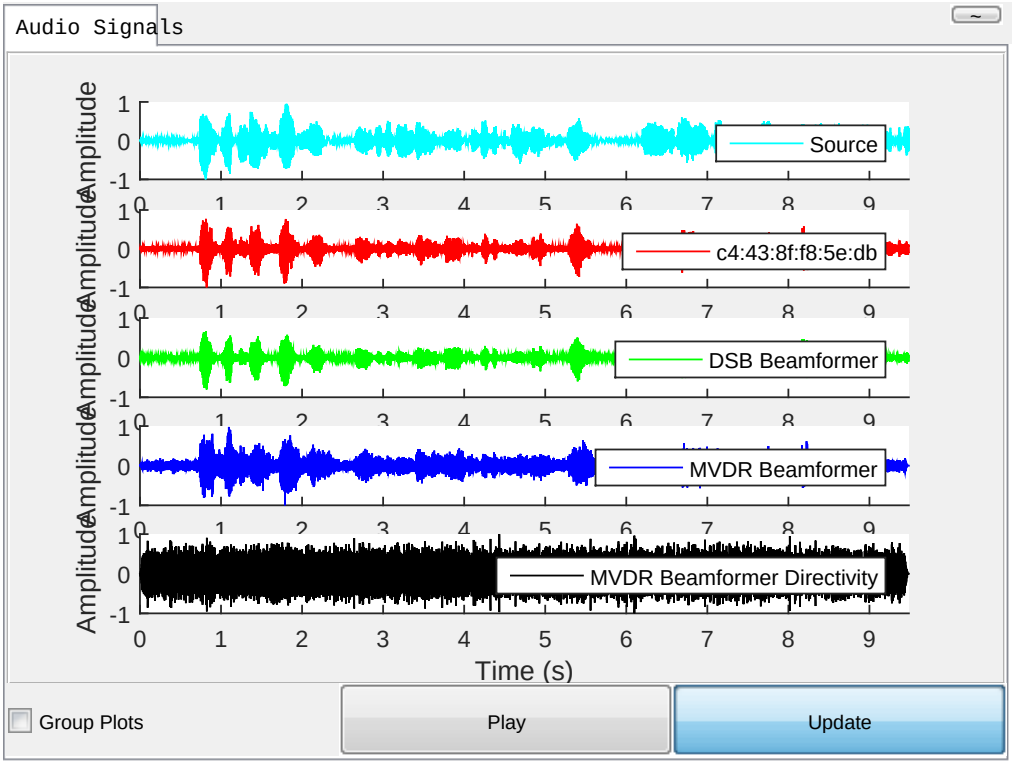


Figure B.15: Audio signals office room experiment 3: With an interfering audio source, $\theta = 90°$

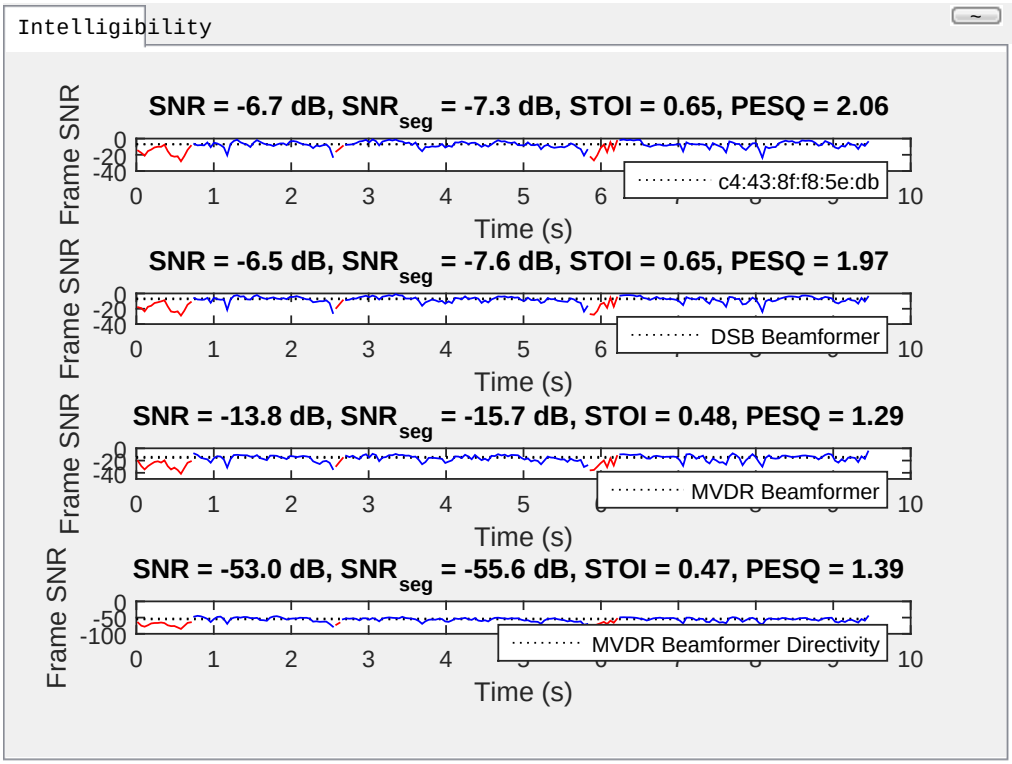


Figure B.16: Intelligibility office room experiment 3: With an interfering audio source, $\theta = 90°$

B.2.4. OFFICE ROOM EXPERIMENT 4

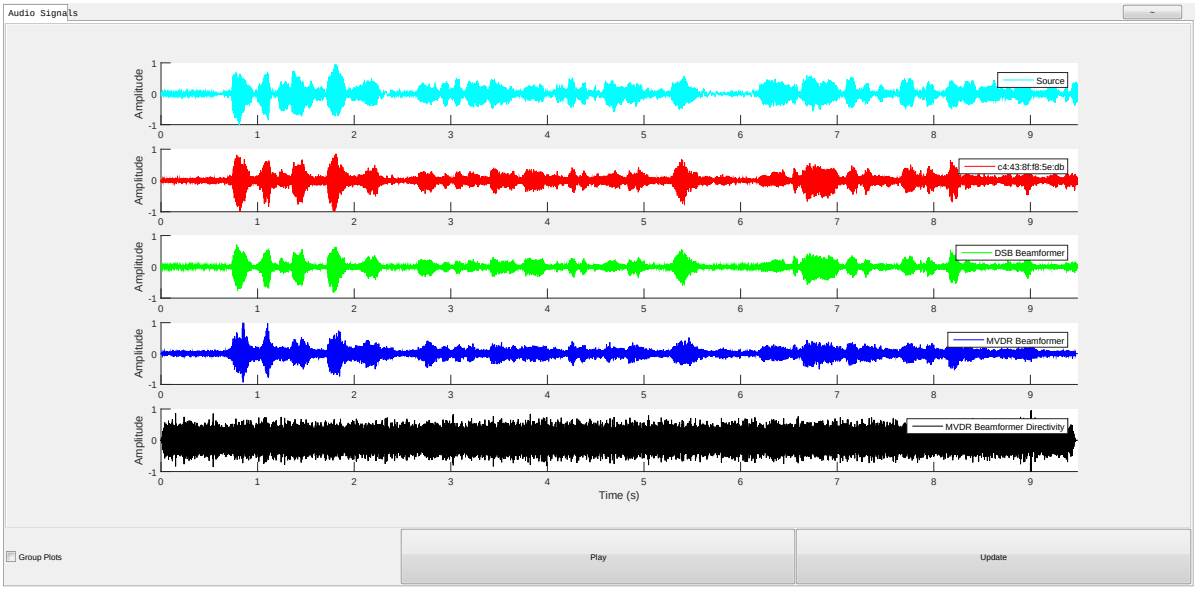


Figure B.17: Audio signals office room experiment 4: With an interfering audio source, Different orientations

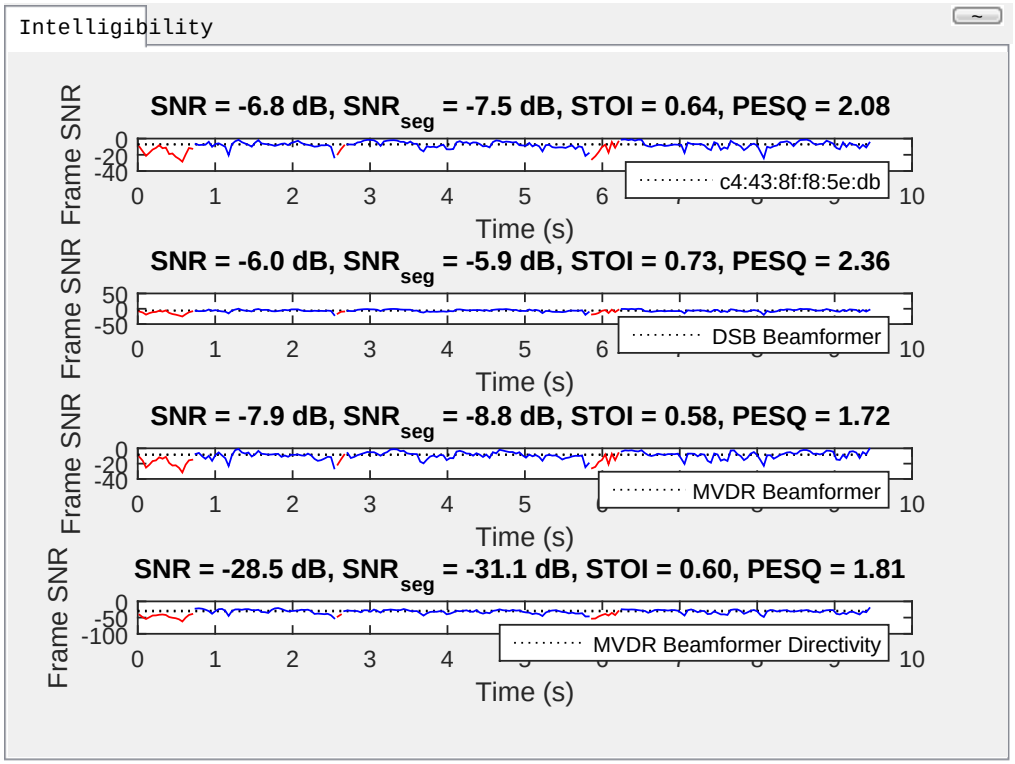


Figure B.18: Intelligibility office room experiment 4: With an interfering audio source, Different orientations

B.2.5. OFFICE ROOM EXPERIMENT 5

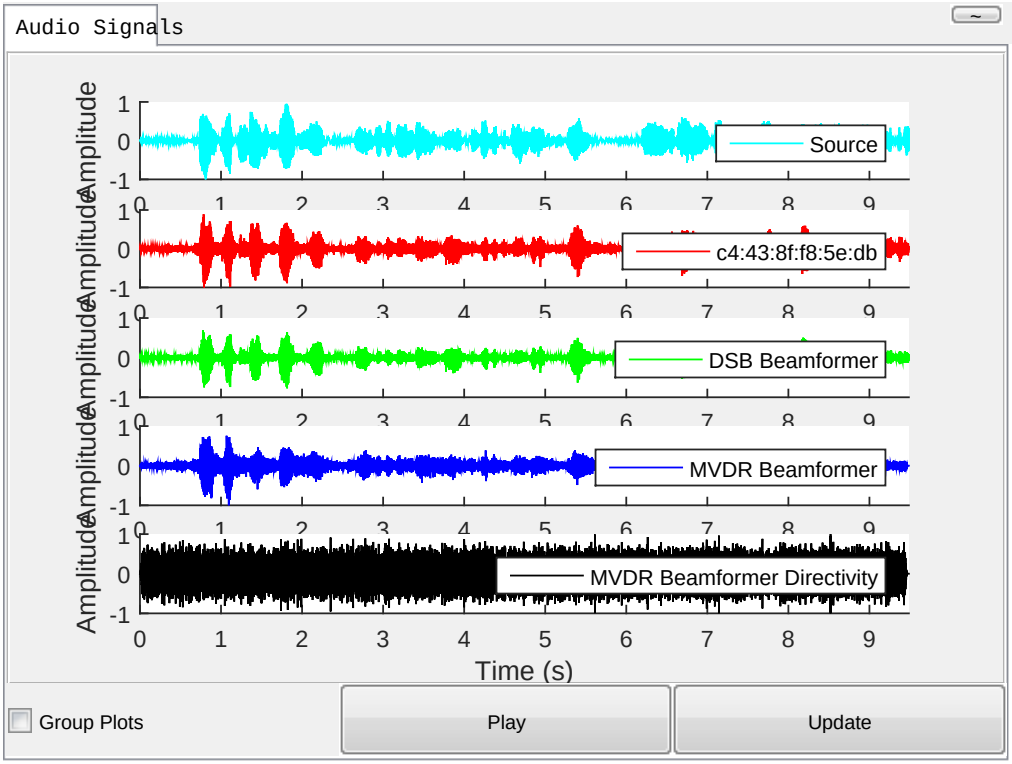


Figure B.19: Audio signals office room experiment 5: With an interfering audio source, Linear array

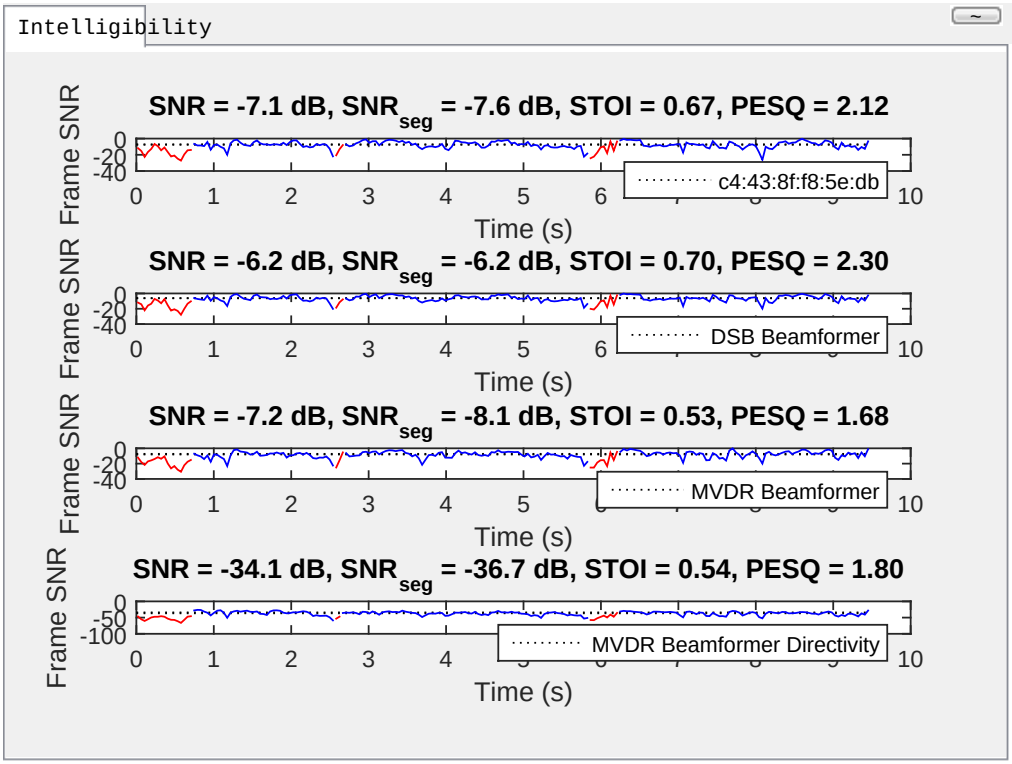


Figure B.20: Intelligibility office room experiment 5: With an interfering audio source, Linear array

B.3. ANECHOIC CHAMBER EXPERIMENTS

B.3.1. ANECHOIC CHAMBER EXPERIMENT 1

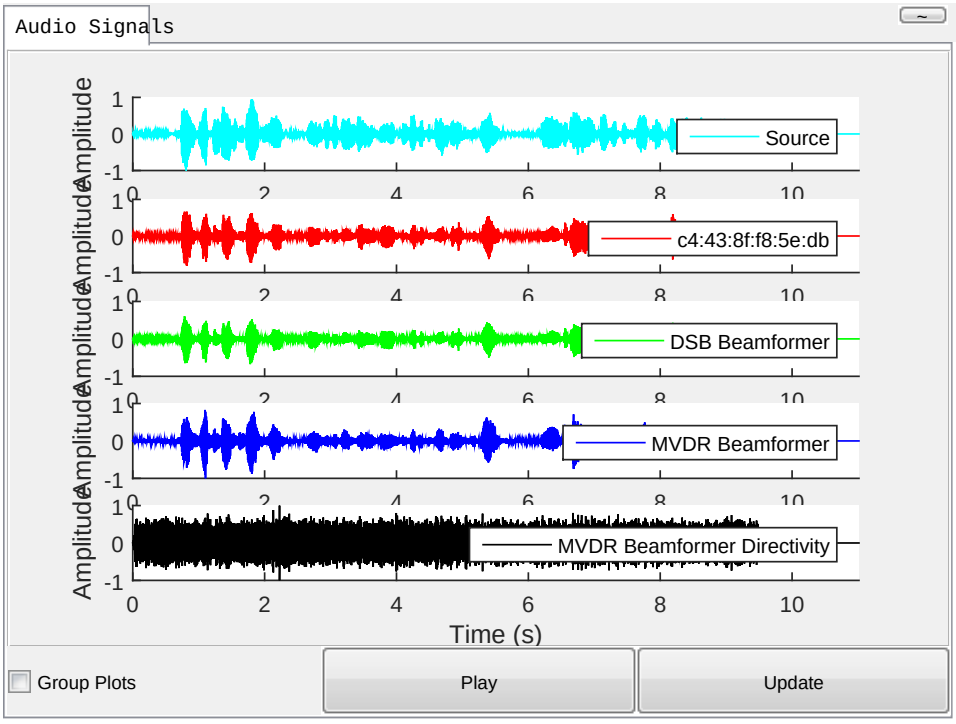


Figure B.21: Audio signals anechoic chamber experiment 1: With an interfering audio source

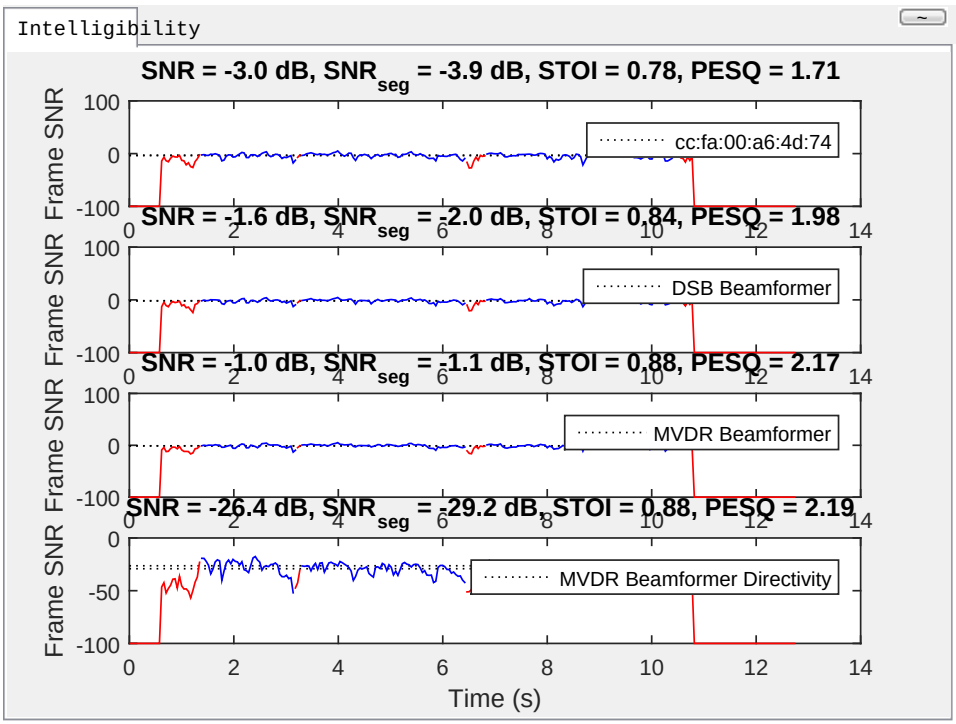


Figure B.22: Intelligibility anechoic chamber experiment 1: With an interfering audio source

BIBLIOGRAPHY

[1] D. Ba, D. Florencio, and C. Zhang, *Enhanced mvdr beamforming for arrays of directional microphones,* in *Multimedia and Expo, 2007 IEEE International Conference on* (2007) pp. 1307–1310.

[2] J. B. Allen and D. A. Berkley, *Image method for efficiently simulating small-room acoustics,* Journal of The Acoustical Society of America **65** (1979), 10.1121/1.382599.

[3] *ITU-T recommendation P.862. Perceptual evaluation of speech quality (PESQ): an objective method for end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,* (2001).

[4] Pentek.com, *Pentek, inc. 8-channel aurora beamforming system - vxs/xmc techcast presentation,* (2015).

[5] B. D. Van Veen and K. M. Buckley, *Beamforming: A versatile approach to spatial filtering.* IEEE ASSP magazine **5**, 4 (1988), cited By 1159.

[6] P. Naylor and N. Gaubitch, *Speech Dereverberation*, Signals and Communication Technology (Springer, 2010).

[7] E. Habets, J. Benesty, S. Gannot, and I. Cohen, *The mvdr beamformer for speech enhancement,* in *Speech Processing in Modern Communication*, Springer Topics in Signal Processing, Vol. 3, edited by I. Cohen, J. Benesty, and S. Gannot (Springer Berlin Heidelberg, 2010) pp. 225–254.

[8] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Digital Signal Processing - Springer-Verlag (Springer, 2001).

[9] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, *Sensitivity analysis of mvdr and mpdr beamformers,* in *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of* (2010) pp. 000416–000420.

[10] I. Himawan, I. McCowan, and S. Sridharan, *Clustered blind beamforming from ad-hoc microphone arrays,* Audio, Speech, and Language Processing, IEEE Transactions on **19**, 661 (2011).

[11] N. Gaubitch, J. Martinez, W. Kleijn, and R. Heusdens, *On near-field beamforming with smartphone-based ad-hoc microphone arrays,* in *Acoustic Signal Enhancement (IWAENC), 2014 14th International Workshop on* (2014) pp. 94–98.

[12] L. J. Griffiths and C. Jim, *An alternative approach to linearly constrained adaptive beamforming,* Antennas and Propagation, IEEE Transactions on **30**, 27 (1982).

[13] A. Kjellson, *Sound source localization and beamforming for teleconferencing solutions,* (2014).

[14] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, *New insights into the mvdr beamformer in room acoustics,* Audio, Speech, and Language Processing, IEEE Transactions on **18**, 158 (2010).

[15] N. G. J. Martinez and W. B. Kleijn, *A robust region-based near-field beamformer in acoustics,* Speech and Signal Processing (ICASSP) **April 2015** (2015).

[16] R. Brinkman and T. de Rooij, *Determining smartphone microphone directivity for application in beamforming*, Bachelor's thesis, Delft University of Technology (2015).

[17] S. Bosma and R. Smeding, *Smartphone Audio Acquisition and Synchronization Using an Acoustic Beacon*, Bachelor's thesis, Delft University of Technology (2015).

[18] M. Gerrits, R. Hendriks, N. Gaubitch, J. Jensen, and M. Pedersen, *Evaluation of instrumental measures for the prediction of musical noise in enhanced noisy speech,* in *3 4 rd Symposium on Information Theory in the Benelux and the 3 rd Joint WIC/IEEE Symposium on Information Theory and Signal Processing in the Benelu x* (2014).

[19] W. Liu and S. Weiss, in *Wideband Beamforming* (John Wiley & Sons, Ltd, 2010) pp. 267–281.

[20] H. Krim and M. Viberg, *Two decades of array signal processing research: the parametric approach,* Signal Processing Magazine, IEEE **13**, 67 (1996).

[21] R. Kennedy, T. Abhayapala, and D. B. Ward, *Broadband nearfield beamforming using a radial beampattern transformation,* Signal Processing, IEEE Transactions on **46**, 2147 (1998).

[22] *ITU-T recommendation P.56. objective measurement of active speech level,* (2011).

[23] tagkey2013225, *Appendix b - op amp noise theory,* in *Op Amps for Everyone (Fourth Edition)*, edited by B. Carter (Newnes, Boston, 2013) fourth edition ed., pp. 225 – 233.

[24] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, *A short-time objective intelligibility measure for time-frequency weighted noisy speech,* in *Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on* (2010) pp. 4214–4217.

[25] A. Rix, J. Beerends, M. Hollier, and A. Hekstra, *Perceptual evaluation of speech quality (pesq)-a new method for speech quality assessment of telephone networks and codecs,* in *Acoustics, Speech, and Signal Processing, 2001. Proceedings. (ICASSP '01). 2001 IEEE International Conference on*, Vol. 2 (2001) pp. 749–752 vol.2.

[26] C. Taal, R. Hendriks, R. Heusdens, and J. Jensen, *An algorithm for intelligibility prediction of time-frequency weighted noisy speech,* Audio, Speech, and Language Processing, IEEE Transactions on **19**, 2125 (2011).

[27] *ITU-T recommendation P.800. methods for subjective determination of transmission quality,* (1998).

[28] *ITU-T recommendation P.830. subjective performance assessment of telephone-band and wideband digital codecs,* (1996).

[29] G. W. Elko, *Microphone array systems for hands-free telecommunication,* Speech Commun. **20**, 229 (1996).

[30] P. M. Peterson, *Simulating the response of multiple microphones to a single acoustic source in a reverberant room,* Journal of The Acoustical Society of America **80** (1986), 10.1121/1.394357.

[31] E. Habets, J. Benesty, I. Cohen, S. Gannot, and J. Dmochowski, *New insights into the mvdr beamformer in room acoustics,* Audio, Speech, and Language Processing, IEEE Transactions on **18**, 158 (2010).

[32] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, *Sensitivity analysis of mvdr and mpdr beamformers,* in *Electrical and Electronics Engineers in Israel (IEEEI), 2010 IEEE 26th Convention of* (2010) pp. 000416–000420.

[33] H. Cox, R. Zeskind, and M. Owen, *Robust adaptive beamforming,* Acoustics, Speech and Signal Processing, IEEE Transactions on **35**, 1365 (1987).

[34] K. Kondo, *Subjective Quality Measurement of Speech: Its Evaluation, Estimation and Applications*, Signals and Communication Technology - Springer-Verlag (Springer, 2012).

[35] J. Hee, *Impulse response measurements using mls,* (2003).

[36] M. Hennecke and G. Fink, *Towards acoustic self-localization of ad hoc smartphone arrays,* in *Hands-free Speech Communication and Microphone Arrays (HSCMA), 2011 Joint Workshop on* (2011) pp. 127–132.

[37] Y. Zeng, R. Hendriks, and R. Heusdens, *Clique-based distributed beamforming for speech enhancement in wireless sensor networks,* in *Signal Processing Conference (EUSIPCO), 2013 Proceedings of the 21st European* (2013) pp. 1–5.

[38] Y. Zeng and R. C. Hendriks, *Distributed delay and sum beamformer for speech enhancement via randomized gossip,* Audio, Speech, and Language Processing, IEEE/ACM Transactions on **22**, 260 (2014).

[39] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn, *Distributed mvdr beamforming for (wireless) microphone networks using message passing,* in *Acoustic Signal Enhancement; Proceedings of IWAENC 2012; International Workshop on* (2012) pp. 1–4.

[40] S. Doclo and M. Moonen, *Design of far-field and near-field broadband beamformers using eigenfilters,* Signal Processing **83**, 2641 (2003).