Graz University of Technology

Master Thesis

Beamforming Using Uniform Circular Arrays for Distant Speech Recognition in Reverberant Environment and Double-Talk Scenarios

conducted at the
Signal Processing and Speech Communication Laboratory
Graz University of Technology, Austria

by
BSc Hannes Pessentheiner, 0573063

Supervisors:
Dipl.-Ing. Dr.sc.ETH Harald Romsdorfer
MSc Dr.techn. Tania Habib

Assessors/Examiners:
Univ.-Prof. Dipl.-Ing. Dr.techn. Gernot Kubin

Graz, February 28, 2012

Statutory Declaration

I declare that I have authored this thesis independently, that I have not used other than the declared sources/resources, and that I have explicitly marked all material which has been quoted either literally or by content from the used sources.

date	(signature)

Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Arbeit selbstständig verfasst, andere als die angegebenen Quellen/Hilfsmittel nicht benutzt, und die den benutzten Quellen wörtlich und inhaltlich entnommene Stellen als solche kenntlich gemacht habe.

Graz, am	(Unterschrift)

Acknowledgement

Foremost, I'ld like to record my gratitude to my supervisor Dipl.-Ing. Dr.sc.ETH Harald Romsdorfer who encouraged me to be not restricted in one's thinking and to spend some time on ideas and procedures of solutions even though they seem to be complex or time-consuming, who supported me with enthusiasm, scientific intuitions, and immense knowledge about speech processing, and who taught me the significance of self-initiated work in order to expand one's knowledge. I gratefully thank MSc Dr.techn. Tania Habib for her guidance and all advices at the very early stage of my thesis. I appreciatively acknowledge Dipl.-Ing. Dr.techn. Stefan Petrik for the deployment of the word recognizer and his patience in introducing me to the world of speech recognition. I am obliged to many of my labmates, especially Antonio Hölzl and Michael Tauch, who supported me all the time with constructive and extensive discussions around my work. My sincere thanks also goes to DI Dr.rer.nat. Franz Zotter and the Institute of Electronic Music and Acoustics for supplying me with additional equipment for my recordings. And it is a pleasure for me to pay tribute to two special people: Ariane Klatzer who is the most important person in my life, who supported me with tons of scientific and personal advices, and who made (and is still making) the world outside of the laboratory much more colorful; and Lukas Hutter who believed in me and my abilities from the beginning of my study, and who got me on the straight and narrow path of music and technology. Finally, I'd like to say that I am truly indebted and thankful to my parents and my brother who made 'the whole thing' possible. Thank you so much!

Abstract

Beamforming is crucial for hands-free mobile terminals and voice-enabled automated home environments based on distant-speech interaction to mitigate causes of system degradation, e.g., interfering noise sources, room reverberation, closed-loop feedback problems, and competing speakers. The objective of this thesis is to find the most common and state-of-the-art broadband beamformers which are able to attenuate or eliminate the competing speaker in case of double-talk scenarios, and which are compatible with the uniform circular microphone array, or—if not—to make them compatible. Moreover, a new beamformer for improved spatial filtering in reverberant environments is introduced. Another objective is to design a MATLAB framework to simplify the implementation of different microphone array geometries and beamformers, and to evaluate their performances and the quality of their corresponding enhanced output signals numerically and graphically by considering different objective measures, e.g., a word recognizer based on a simple grammar and a limited dictionary that covers all words appearing in the CHiME-Corpus and audio signals used in this work. For the evaluation, speech signals are played-back synchronously and separately by two loudspeakers in a reverberant environment, recorded by a uniform circular microphone array, and subsequently filtered by different beamformers.

Zusammenfassung

Heutzutage ist Beamforming ein wichtiger Bestandteil im Bereich der Telekommunikation und Sprachsteuerung, um Störeinflüsse wie unerwünschte Rauschquellen, konkurrierende Sprecher, Nachhall oder Rückkopplungsschleifen zu unterdrücken. Ziele dieser Arbeit sind das Finden von Beamformern, die mit einem kreisförmigen Mikrofon-Array kompatibel sind, das Anpassen von nicht kompatiblen Beamformern, und die Entwicklung eines neuen Beamformers zur besseren Unterdrückung des konkurrierenden Sprechers im Fall eines Double-Talk–Szenarios. Ein weiteres Ziel dieser Arbeit ist die Erstellung einer Simulations- und Auswertungsumgebung in MATLAB zur einfachen Einbindung verschiedener Mikrofon-Array Geometrien und Beamformern, und zur grafischen und numerischen Qualitätsbeurteilung von Beamformern und den von ihnen gefilterten Signalen. Neben den bekannten Beurteilungsmaßen für die gefilterten Signale findet auch ein Spracherkenner Verwendung, welcher auf einer einfachen Grammatik basiert und eine bestimmte Anzahl verschiedener Wörter erkennt, die im CHiME-Korpus definiert und in den verwendeten Audio-Signalen vorhanden sind. Für die Evaluierung der Beamformer-Performance und der Qualität der gefilterten Signale wurden bestimmte Sprachsignale mittels zweier Lautsprecher in einem halligen Raum ausgegeben, mit einem kreisförmigen Mikrofon-Array aufgenommen und anschließend mit den vorhandenen Beamformern gefiltert.

Contents

1 Introduction

1.1 Introduction

During World War I scientists realized that using antenna arrays for secret wireless communication entails directional transmission of information under certain circumstances. Armed forces exploited this knowledge in World War II to communicate with allies without sending their information radially (in all directions). Consequently, opposing forces were not able to easily intercept secret messages at any position in the immediate vicinity of the arrays. To improve the quality of long distance communication, military intelligence services used these arrays for message reception to eliminate disturbing interferences produced by opposing forces, natural noise sources, or atmospheric disturbances.

Nowadays, the use of antenna arrays—in general: sensor arrays—is a fundamental and important technique to improve data transmission or data reception over long distances, e.g., interplanetary communication between satellites, radio astronomy, etc. In everyday life, sensor arrays improve communication between, e.g., underwater research facilities and their corresponding submarine vehicles, or hand-held devices and mobile phone base stations. Especially, in times of hands-free functionality of mobile devices, microphone arrays become more and more relevant as the following scenario shows: a person is driving a car and wants to phone a friend without using its hands; therefore, it's necessary to use the mobile device in hands-free mode. The mobile phone is generally fixed near the instrument panel or somewhere at the dashboard. Phoning without any modifications in hardware or without any additional signal processing techniques for speech enhancement and noise reduction leads to distorted and noisy communication because of noise and interferences produced by mechanical vibrations of the wheels and the engine of the car [1]. A way around the problem is to consider beamforming in order to attenuate noise and interferences—both arrive from all directions—and to set a focus on the car-driving speaker. A prerequisite for using beamforming techniques is to equip the hand-held devices with multiple microphones (omnidirectional or directional) and proper signal processing techniques, e.g., source localization or source tracking, source separation, and beamforming algorithms.

1.2 Motivation

In audio signal processing beamforming provides the ability to separate two or more sound sources. It is used as an acoustic camera to focus 'desired' and eliminate 'interfering' sources. The choice of the right beamformer depends on the operating environment or working area, e.g., a reverberant conference room, home environments, etc. Beamforming eliminates causes of system degradation—interfering noise sources, room reverberation, closed-loop feedback problems, and competing speakers—in case of full-duplex teleconferencing. It is fundamental for hands-free mobile terminals [2] (see Fig. 1.1) and for voice-enabled automated home environments based on distant speech interaction, where a distributed microphone network enables the monitoring of speech activity within a room (see DIRHA[1]). Furthermore, the right choice of the corresponding microphone array is as important as the right choice of the beamformer. The UCA[2] increases the performance of the beamformer when the source distance is not known, and it enables focusing sources which are larger than the microphone array.

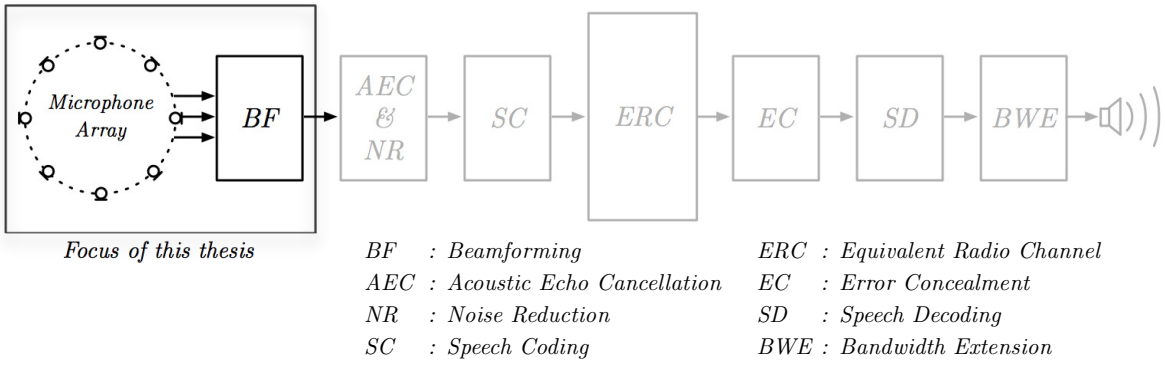


Figure 1.1: This figure shows the speech signal processing in a hands-free mobile terminal. The boxed elements highlight the focus of this thesis.

1.3 Objective

The objective of this work is to find the most common, state-of-the-art broadband beamformers which are compatible with the UCA, or—if not—to make them compatible, modify an existing, or introduce a new beamformer for improved spatial filtering in reverberant environments and double-talk scenarios. Another objective is to design a MATLAB framework to simplify the implementation of different microphone array geometries and beamformers, and to evaluate the beamformers' performance and the quality of their corresponding enhanced output signals numerically and graphically by considering different objective measures and a word recognizer based on a simple grammar and a limited dictionary that covers all words appearing in the CHiME-Corpus and audio signals used in this work. For the evaluation, specially composed signals are played-back by loudspeakers in a reverberant environment, recorded by a UCA, and subsequently filtered by different beamformers.

[1] DIRHA - Distant Speech Interaction for Robust Home Applications: `http://dirha.fbk.eu/`
[2] UCA - Uniform Circular Array

2 Sound Capture with Microphone Arrays

2.1 Capturing Sound

It is hard to cope with problems like reverberation, noise, and multiple sound sources in two- or three-dimensional sound propagation processes, especially, when a single microphone is available only [3]. The use of multiple microphones offers directional gains, which improves the signal quality and enables source focusing, noise and interference attenuation; furthermore, it enables the estimation of the TDOA[3], which is fundamental for source localization and beamforming [3]. The microphones span a region where they capture information in terms of audio signals. This information is generally considered as energy. The energy receiving region is the aperture, which can be either discretely or, theoretically, continuously realized [4]. A discrete realization enables signal processing on each channel, but it introduces spatial aliasing and grating lobes; both are discussed in detail in Section 3.1.3 and Section 3.2.3. The microphones can be placed in different ways as shown in Fig. 2.1: linearly, circularly, rectangularly, uniformly, non-uniformly, etc. Each array geometry exhibits its advantages and drawbacks, e.g., on the one hand a ULA[4] based on omnidirectional microphones is easy to use and narrowband operations work efficiently, on the other hand it has to cope with the front-back ambiguity, and the spatial aliasing frequency[5] strongly depends on the steering direction ϕ_s, also known as the steering angle, focusing direction or main response axis. It is the angle of the position of the desired source provided by a source localization algorithm. In this work, ϕ_s is an a priori knowledge. In contrast to linear arrays, planar arrays eliminate the front-back ambiguity, and they reduce the angle-dependency of the spatial aliasing frequency, but complexity in signal processing increases.

[3] TDOA - Time Delay of Arrival
[4] ULA - Uniform Linear Array
[5] It is the frequency where a side lobe appears in the array response or beam pattern.

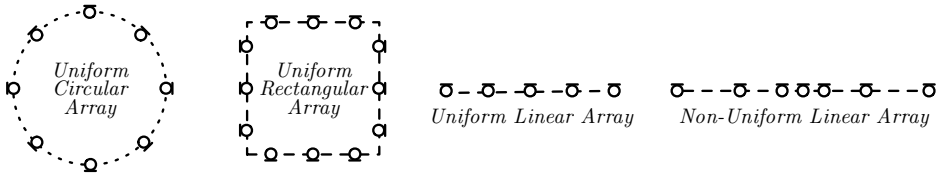


Figure 2.1: This figure shows four different types of microphone arrays. The UCA and the uniform rectangular array are planar arrays, whereas the uniform and non-uniform linear array are referred to as line arrays.

2.1.1 The Wave Propagation

In order to understand how microphone arrays operate, it is necessary to introduce the basic concept of planar wave propagation. It enables the estimation of a sound field by measuring acoustic parameters within a certain area, e.g., acoustic pressure $p(x, y, z, t)$ measured by microphones of a microphone array [5], which is an important parameter in the modeling of acoustic wave propagations. The acoustic wave equation

$$\nabla^2 p(x, y, z, t) = \frac{1}{c^2} \frac{\partial^2 p(x, y, z, t)}{\partial t^2}, \tag{2.1}$$

derived by Richard Feynman, describes a simple linear propagation model [4][5][6]. In this model p is the instantaneous sound pressure fluctuation, and c is the sound velocity or the propagation speed of sound. The sound pressure depends on three space variables (x, y, z) and one time variable (t). In simple terms, $p(\boldsymbol{r}, t)$ represents the acoustic pressure field, where $\boldsymbol{r} = (x, y, z)^T$ describes the position of a microphone. The four-dimensional Fourier transform, in this case the temporal (one-dimensional) and the spatial (three-dimensional) Fourier transform of $p(\boldsymbol{r}, t)$, results in

$$P(\boldsymbol{k}, \omega) = \int_{-\infty}^{+\infty} \left(\int_{-\infty}^{+\infty} p(\boldsymbol{r}, t) e^{-i\omega t} dt \right) e^{\boldsymbol{k}^T \boldsymbol{r}} d\boldsymbol{r} = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} p(\boldsymbol{r}, t) e^{-i(\omega t - \boldsymbol{k}^T \boldsymbol{r})} d\boldsymbol{r} dt$$

and its inverse

$$p(\boldsymbol{r}, t) = \frac{1}{(2\pi)^4} \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} P(\boldsymbol{k}, \omega) e^{i(\omega t - \boldsymbol{k}^T \boldsymbol{r})} d\boldsymbol{k} d\omega, \tag{2.2}$$

where $\boldsymbol{k}$ is the wavevector

$$\boldsymbol{k} = -k \begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix} = -\frac{2\pi}{\lambda} \begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix} = -\frac{2\pi f}{c} \begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix} = -\frac{\omega}{c} \begin{pmatrix} k_x \\ k_y \\ k_z \end{pmatrix} = -\frac{\omega}{c} \begin{pmatrix} \sin(\theta)\cos(\phi) \\ \sin(\theta)\sin(\phi) \\ \cos(\theta) \end{pmatrix},$$

which depends on the angular frequency ω, the angular wavenumber k ($|\boldsymbol{k}| = 2\pi/\lambda$ is the magnitude), and the directional information (k_x, k_y, k_z) retrieved from the spherical coordinates. The variables θ and φ represent the elevation and azimuth as shown in Fig. 2.2. In this work the elevation θ is set to 90°. The wavevector describes the phase variation of a monochromatic plane wave, and its components k_x, k_y and k_z define the change of phase in the corresponding direction. In simple words, the wavevector gives information about the direction of propagation and the wavelength of the monochromatic plane wave. In real scenarios a microphone captures an infinite number of monochromatic waves from all directions at each point of time, e.g.,

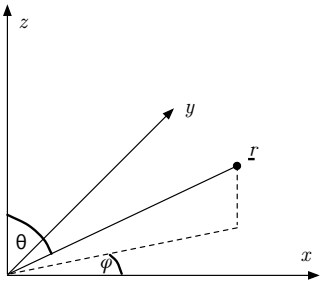


Figure 2.2: The coordinate system used in this work with its azimuth ϕ, its elevation θ, and its position-vector $\boldsymbol{r}$,e.g., a microphone position, and its coordinates (x,y,z).

reflections, noise sources, etc., according to

$$s_0(\boldsymbol{r}_0, t) = \sum_{j=0}^{\infty} A_j e^{i(\omega_j t - \boldsymbol{k}_j^T \boldsymbol{r}_0)} = p(\boldsymbol{r}_0, t)$$

where $p(\boldsymbol{r}_0, t)$ is the sound pressure field at position $\boldsymbol{r}_0$, and $s_0(\boldsymbol{r}_0, t)$ is the captured signal with a microphone on position $\boldsymbol{r}_0$. In the theoretical part of this work $s_0(\boldsymbol{r}_0, t)$ is modeled as a single, monochromatic plane wave. Thus, $s_0(\boldsymbol{r}_0, t)$ assumes an anechoic room with a single source:

$$s_0(\boldsymbol{r}_0, t) = A_0 e^{i(\omega_0 t - \boldsymbol{k}_0^T \boldsymbol{r}_0)}.$$

The Near-Field Model

The near-field is the immediate area around a source. The sound pressure and the sound particle velocity are not in phase. The captured signal of a microphone is

$$S_n(\omega, \boldsymbol{r}_n, \boldsymbol{r}_s) = D_n(\omega, \boldsymbol{r}_n, \boldsymbol{r}_s) \cdot R(\omega) + N_n(\omega)$$

with

$$D_n(\omega, \boldsymbol{r}_n, \boldsymbol{r}_s) = \frac{1}{||\boldsymbol{r}_s - \boldsymbol{r}_n||} A_n(\omega) U_n(\omega) e^{-i\frac{\omega}{c}||\boldsymbol{r}_s - \boldsymbol{r}_n||},$$

where $D_n(\omega, \boldsymbol{r}_n, \boldsymbol{r}_s)$ is the sound capture model of a microphone with index n, $||\boldsymbol{r}_s - \boldsymbol{r}_n||$ is the distance between the source and the microphone with index n, $A_n(\omega)$ is the frequency response of an amplifier and/or an ADC[6], $U_n(\omega)$ represents the microphone characteristics, $e^{-i(\cdot)}$ describes the phase rotation due to the distance between the microphone and the source, $R(\omega)$[7] is the source signal in frequency domain, and $N_n(\omega)$ is the noise modeled as a zero-mean Gaussian random process in frequency domain. A near-field model is physically more precise than a far-field model, and it is valid for both fields, the near- and far-field, but it requires the source direction and the distance between the source and the microphone array. It is better to use the near-field model in case of numerical approximations [7] of the optimal beamformer coefficients.

[6] ADC - Analog to Digital Converter

[7] Do not mix up $R(\omega)$ with the acoustic pressure field $P(\boldsymbol{k}, \omega)$. $R(\omega)$ does not contain any directional information.

The Far-Field Model

The far-field is—as its name implies—far away from the source. The sound pressure and the sound particle velocity are in phase. The captured signal of a microphone is

$$S_n(\omega, \boldsymbol{r}_n, \boldsymbol{\eta}) = D_n(\omega, \boldsymbol{r}_n, \boldsymbol{\eta}) \cdot R(\omega) + N_n(\omega)$$

with

$$D_n(\omega, \boldsymbol{r}_n, \boldsymbol{\eta}) = \frac{1}{\rho} A_n(\omega) U_n(\omega) e^{-i\frac{\omega}{c}||\boldsymbol{r}_n|| \cos(\phi_s - \phi_n)},$$

where $\boldsymbol{\eta} = (\rho, \phi_s, \phi_n)^T$, ρ is the average distance between the source and each microphone, i.e. the attenuation $1/\rho$ of the source signal captured by each microphone is the same. It is easier to do calculations with the far-field model, and it is more suitable for determining the weighting coefficients analytically. It depends on the source direction only.

2.1.2 The Wavevector-Frequency Domain

Equation (2.2) shows that an infinite number of complex weighted propagating monochromatic plane waves can reconstruct the sound pressure field $p(\boldsymbol{r}, t)$ under consideration that the monochromatic plane wave in the spatio-temporal domain is given as

$$p_0(\boldsymbol{r}, t) = A_0 e^{i(\omega_0 t - \boldsymbol{k_0}^T \boldsymbol{r})}, \tag{2.3}$$

where $\boldsymbol{k_0}$ defines the direction of propagation. According to [4] the wavevector-frequency representation of (2.3) is

$$P_0(\boldsymbol{k}, \omega) = A_0 \cdot (2\pi)^4 \cdot \delta(\boldsymbol{k} - \boldsymbol{k_0}) \cdot \delta(\omega - \omega_0),$$

which yields a single point in the wavevector-frequency space spanned by the spatial-frequency and the temporal-frequency space due to the one- and three-dimensional Dirac-impulse. Fig. 2.3 depicts a wavevector-frequency space for three different scenarios. The distance between the center of the coordinate system and a single point defines the wave length or frequency of the wave, the brightness of a point describes the magnitude, and the position unveils information about the direction of propagation.

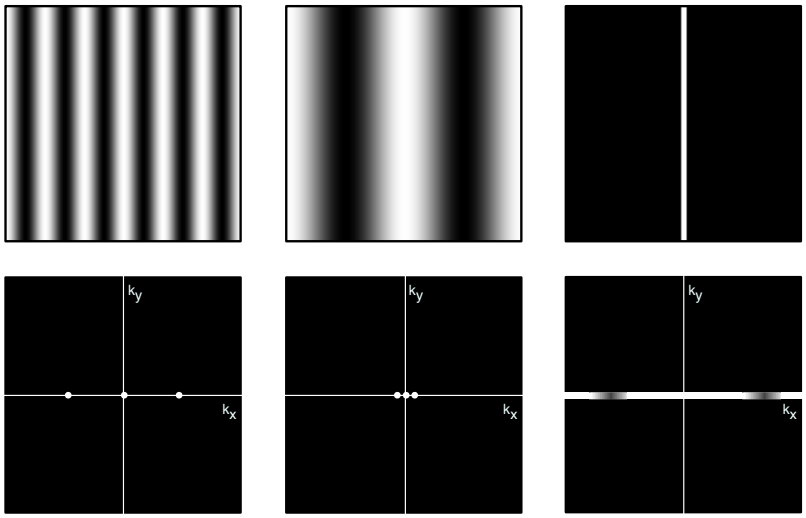
An important Fourier property in temporal-frequency domain is the representation of a convolution in time domain as a multiplication in frequency domain; and so it is in the spatial-frequency domain. Thus, a convolution in spatio-temporal domain

$$y(\boldsymbol{r}, t) = \int_{-\infty}^{+\infty} \int_{-\infty}^{+\infty} h(\boldsymbol{r} - \boldsymbol{\tau_1}, t - \tau_2) p(\boldsymbol{\tau_1}, \tau_2) d\boldsymbol{\tau_1} d\tau_2$$

corresponds to

$$Y(\boldsymbol{k}, \omega) = H(\boldsymbol{k}, \omega) \cdot P(\boldsymbol{k}, \omega),$$

in wavevector-frequency domain, where $h(\boldsymbol{r}, t)$ is the spatio-temporal impulse response, and $H(\boldsymbol{k}, \omega)$ is its wavevector-frequency representation. This property enables filtering the acoustic scalar pressure field in wavevector-frequency domain, which is exploited by *beamforming*. The following equation manipulates the frequency response of propagating waves from a given



A sinusoidal wave with high frequency propagates on the k_x-axis.

A sinusoidal wave with low frequency propagates on the k_x-axis.

An impulse propagates on the k_x-axis.

Figure 2.3: Visualization of a plane wave in the wavevector-frequency domain.

direction $\boldsymbol{k}_0$

$$H(\boldsymbol{k}, \omega) = \delta(\boldsymbol{k} - \boldsymbol{k_0}) \cdot G(\omega),$$

where $G(\omega)$ is the frequency response and $\boldsymbol{k_0}$ describes the direction of the propagating wave.

2.1.3 Beamforming with Microphone Arrays

The main idea of beamforming with microphone arrays is to set a focus in a certain direction—the steering direction ϕ_s—with the result that signals from this direction captured by the microphones overlap constructively; signals from other directions overlap deconstructively: they are attenuated. Typical signals are speech or music, both are broadband signals. A continuous realization of an array with microphones or microphone capsules is not possible; thus, a discretized microphone array samples the acoustic field within a specified region: the aperture. The number of microphones equals the number of channels and spatial Fourier transform kernels. For instance, the DS-BF[8] summarizes the time-shifted signals captured by all microphones and normalizes the sum with the number of microphones according to [7]

$$y(t) = \frac{1}{N} \sum_{n=1}^{N} s(\boldsymbol{r}_n, t - \tau_n),$$

where N is the number of microphones, $\boldsymbol{r}_n$ represents the position of the microphone with index n, τ_n is a microphone-position and steering-direction specific delay which yields constructive overlapping for signals from the direction ϕ_s, and $y(t)$ is the mono-output of the beamformer. In this work the signal captured by each microphone is modeled as a monochromatic plane wave:

$$s(\boldsymbol{r}_n, t) = A_0 e^{i(\omega_0 t - \boldsymbol{k}_0^T \boldsymbol{r}_n)}.$$

[8] DS-BF - Delay-and-Sum Beamformer

Summarizing all captured waves without considering a beamformer leads to

$$y(t) = \sum_{n=1}^{N} A_0 e^{i(\omega_0 t - \boldsymbol{k}_0^T \boldsymbol{r}_n)}.$$

This special case yields constructive overlapping only if all microphones are placed on a straight line symmetrically around the x-axis—$\boldsymbol{r}_n = (0, y_n, 0)^T$ (see Fig. 2.4)—by assuming that a monochromatic plane wave propagates from $\phi_s = 0°$, i.e. $\boldsymbol{k} = (k_x, 0, 0)^T$ and $\boldsymbol{k}^T \boldsymbol{r}_n = 0$. A change in direction of the propagating wave of about 45° without changing the steering direction ϕ_s yields a constructive interference for just a few frequencies within a broadband spectrum. Thus, the sum of all signal components of a periodic broadband signal does not overlap constructively[9] for all directions except 0° and ±180°. Let's consider the following example: Two microphones ($N = 2$) are placed symmetrically around the x-axis. The distance between each microphone is $d = 0.05$ m.

$$y(t) = \sum_{n=1}^{2} A_0 e^{i(\omega_0 t - \boldsymbol{k}^T \boldsymbol{r}_n)} \tag{2.4}$$

The wavevector $\boldsymbol{k}_0$ and the microphone position vector $\boldsymbol{r}_n$ can be rewritten as

$$\boldsymbol{k}_0 = -\frac{\omega_0}{c} \begin{pmatrix} \sin(\theta_s)\cos(\phi_s) \\ \sin(\theta_s)\sin(\phi_s) \\ \cos(\theta_s) \end{pmatrix} = -\frac{\omega_0}{c} \begin{pmatrix} \cos(\phi_s) \\ \sin(\phi_s) \\ 0 \end{pmatrix},$$
$$\boldsymbol{r}_n = r_n \begin{pmatrix} \sin(\theta_n)\cos(\phi_n) \\ \sin(\theta_n)\sin(\phi_n) \\ \cos(\theta_n) \end{pmatrix} = r_n \begin{pmatrix} \cos(\phi_n) \\ \sin(\phi_n) \\ 0 \end{pmatrix},$$

where $\theta_s = 90°$ is the elevation of the source and $\theta_n = 90°$ is the elevation of a microphone with index n. This work assumes an array where all microphones are placed on the xy-plane with $\theta_n = 90°$. Multiplying both vectors in terms of a scalar product yields

$$\boldsymbol{k}_0^T \boldsymbol{r}_n = -\frac{\omega_0}{c} r_n \left(\cos(\phi_s)\cos(\phi_n) + \sin(\phi_s)\sin(\phi_n)\right) = -\frac{\omega_0}{c} r_n \cos(\phi_s - \phi_n),$$

Hence, (2.4) can be rewritten as

$$y(t) = \sum_{n=1}^{2} A_0 e^{i(\omega_0 t + \frac{\omega_0}{c} r_n \cos(\phi_s - \phi_n))} = A_0 e^{i\omega_0 t} \sum_{n=1}^{2} \underbrace{e^{i\frac{\omega_0}{c}\frac{d}{2}\cos(\phi_s - \phi_n)}}_{D_n(\omega_0, \phi_s)}$$
$$= A_0 e^{i\omega_0 t} \left(e^{i\frac{\omega_0}{c}\frac{d}{2}\cos(\phi_s - \phi_1)} + e^{i\frac{\omega_0}{c}\frac{d}{2}\cos(\phi_s - \phi_2)} \right)$$

for $r_n = |\pm\frac{d}{2}| = \frac{d}{2}$ (the distance $\boldsymbol{r}_n$ between the center of the coordinate system and the microphone with index n has to be positive), and $D_n(\omega, \phi_s)$ is the sound capture model of a microphone with index n. Assuming that $\phi_1 = +90°$ and $\phi_2 = -90°$ yields

$$y(t) = A_0 e^{i\omega_0 t} \left(e^{i\frac{\omega_0}{c}\frac{d}{2}\cos(\phi_s - 90°)} + e^{i\frac{\omega_0}{c}\frac{d}{2}\cos(\phi_s + 90°)} \right) = A_0 e^{i\omega_0 t} \left(e^{i\frac{\omega_0}{c}\frac{d}{2}\sin(\phi_s)} + e^{-i\frac{\omega_0}{c}\frac{d}{2}\sin(\phi_s)} \right)$$
$$= A_0 e^{i\omega_0 t} \frac{2}{2} \left(e^{i\frac{\omega_0}{c}\frac{d}{2}\sin(\phi_s)} + e^{-i\frac{\omega_0}{c}\frac{d}{2}\sin(\phi_s)} \right) = A_0 e^{i\omega_0 t} \cdot 2\cos\left(\frac{\omega_0}{c}\frac{d}{2}\sin(\phi_s) \right).$$

[9] In this work captured signals overlap constructively if all frequency-components of the superposition of these signals exhibit constructive interference (perfect fit).

For an even number of microphones and an aperture mentioned above the output—the sum of all signals—is

$$\boxed{y(t) = A_0 e^{i\omega_0 t} \cdot 2 \left(\sum_{k=1}^{N/2} \cos \left[\frac{\omega_0}{c} (2k-1) \frac{d}{2} \sin(\phi_s) \right] \right)},$$

and the output for an odd number of microphones with a microphone in the middle of the coordinate system, the same microphone spacing, and the same array alignment as mentioned before (see Fig. 2.4) is

$$\boxed{y(t) = A_0 e^{i\omega_0 t} \cdot 2 \left(1 + 2 \sum_{k=1}^{(N-1)/2} \cos \left[\frac{\omega_0}{c} (2k-1) \frac{d}{2} \sin(\phi_s) \right] \right)}. \tag{2.5}$$

One can see that the amplitude of the output signal depends on the amplitude of the wave A_0, its frequency ω_0, its direction of propagation ϕ_s, the microphone spacing d, the sound velocity c, and the number of microphones N. The main task of a DS-BF is to align—delay or advance—the signals, so that signals from ϕ_s and captured by the microphones at position $\boldsymbol{r}_n$ overlap constructively. This yields

$$y(t) = \frac{1}{N} \sum_{n=1}^{N} A_0 e^{i(\omega_0 [t-\tau_n] - \boldsymbol{k}_0^T \boldsymbol{r}_n)} = A_0 e^{i\omega_0 t} \cdot \frac{1}{N} \sum_{n=1}^{N} e^{-i(\omega_0 \tau_n + \boldsymbol{k}_0^T \boldsymbol{r}_n)}. \tag{2.6}$$

All delays τ_n in the exponent of (2.6) have to be determined in a way that both terms $\omega_0 \tau_n$ and $\boldsymbol{k}_0^T \boldsymbol{r}_n$ cancel out each other for the steering direction ϕ_s. Signals from different directions overlap destructively[10]. It is noteworthy that the exponent in (2.6) gets positive for negative $\boldsymbol{k}_0^T \boldsymbol{r}_n$ and $|\boldsymbol{k}_0^T \boldsymbol{r}_n| > |\omega_0 \tau_n|$, i.e. the system exhibits non-causal behaviour. An additional delay T_0 eliminates the non-causality [4], which results in

$$y(t) = A_0 e^{i\omega_0 t} \cdot \frac{1}{N} \sum_{n=1}^{N} e^{-i(\omega_0 [\tau_n + T_0] + \boldsymbol{k}_0^T \boldsymbol{r}_n)}. \tag{2.7}$$

Microphone arrays provide the ability to increase the quality of the captured sound. In general, microphone arrays are better than a single microphone, because they increase the SNR[11] of the captured signals. The more elements are used without changing the distance between the elements, the better the array works at higher frequencies because of an increase of the grating lobe frequency f_{gl}. If the sensor spacing decreases, the spatial aliasing frequency f_{sa} and the grating lobe frequency f_{gl} increases. If the array consists of fewer elements, it becomes sensitive to noise, reverberation, and other interferences at higher frequencies.

The use of microphone arrays also introduces some problems which limit the performance:

- inherent noise of the microphones,
- deviations in the microphone frequency responses, i.e. manufacturing variations, and
- deviations in the microphone positions.

These problems—they introduce channel mismatches (see Section 2.2)—require robust micro-

[10] In this work captured signals overlap deconstructively if one or more frequency-components of the superposition of these signals exhibit deconstructive interference.

[11] SNR - Signal to Noise Ratio

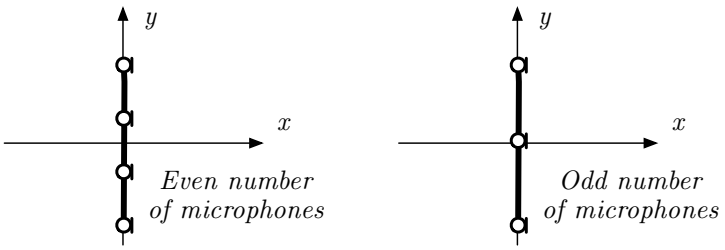


Figure 2.4: The left figure shows a ULA which consists of an even number of microphones, whereas the right one consists of an odd number of microphones.

phone array signal processing algorithms, e.g. a MPDR-BF[12] with proper loading level, i.e. a standard MPDR-BF with additional constraints (see Section 4.4).

2.2 Channel Mismatch

Identical microphones, preamplifiers, and ADCs are necessary for perfect channel matching [7]; they do not exhibit any deviations in their characteristics. In general, the sources of deviations—they evoke channel mismatches and, thus, performance losses—are the manufacturing tolerances of microphones and microphone arrays. These manufacturing tolerances affect the sensitivity, the magnitude, the phase, and the frequency responses towards the main response axis or steering direction ϕ_s.

The microphone and preamplifier parameters alter due to

- variation in temperature,
- variation in atmospheric pressure, and
- disturbances in the power supply.

The microphone array aperture additionally introduces deviations because of inaccurate drillings. All these deviations influence the output of the beamformer negatively, which results in lower performances of the array signal processing algorithms. Thus, ideal array signal processing requires precise manufacturing and stationary processes, which is difficult to accomplish. Robust algorithms reduce the mismatch-sensitivity.

A real array response for a certain direction may look as follows:

$$\boxed{U(f) = U_{opt}(f)M(f)e^{-i\varphi(f)}}\,,$$

where $U_{opt}(f)$ represents the optimal microphone characteristic, $M(f)$ represents the effects of variations in sensitivity modeled as a normal distribution $\mathcal{N}(1, \sigma_{M(f)})$, and $\varphi(f)$ represents the phase-deviations modeled as $\mathcal{N}(0, \sigma_{\varphi(f)})$. According to [7] the influence of deviations in magnitudes is much higher than the influence of variations in phase. An important task in array signal processing is to eliminate/compensate these deviations by considering, e.g., calibration.

[12] MPDR-BF - Minimum Power Distortionless Response Beamformer

2.3 Gain Self-Calibrating Algorithms

Calibrating algorithms mitigate the mismatch in magnitude [7]. There are three different types of calibration:

- the pre-design calibration,
- the post-installation calibration, and
- the self-calibration.

The pre-design calibration requires measurements of the directivity pattern of each microphone channel. Individual filters compensate the magnitude mismatch for each channel separately before beamforming.

The post-installation calibration considers the calibration of the microphone array at a certain position in a room. The location of the source position has to be known in advance. The calibration is done by using white-noise calibration signals in far-field condition and by computing the filter coefficients with the NLMS[13]-method or similar adaptive methods. This calibration method requires manual calibration after the setup of the microphone array.

The third and most convenient calibration method is the gain self-calibration based on real-time computations. Consequently, it requires a certain amount of CPU time to eliminate gain-mismatches adaptively. There are two different methods to calibrate the channels: One method is active during pauses only, the other one during speech. The first one assumes omnidirectional microphones, far-field condition, and a compact microphone array, i.e. it should exhibit a small diameter. If all assumptions are fulfilled, all signals captured by the microphones exhibit approximately the same sound level, which is fundamental for the following calculations:

$$G_m = \frac{\bar{L}_n}{L_{m,n}},$$

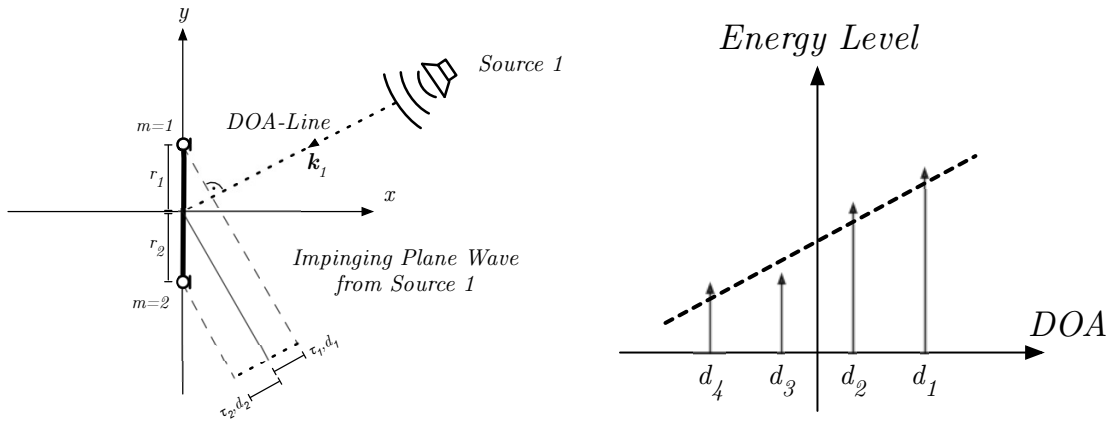
where G_m is the individual gain of channel m, $\bar{L}$ is the averaged RMS[14] over all channels, L_m is the RMS of channel m, and n is the temporal frame number. Smoothing may increase the performance of the calibration algorithm:

$$G_{m,n} = \left(1 - \frac{T}{\tau_G}\right) G_{m,n-1} + \frac{T}{\tau_G} G_m,$$

where $G_{m,n}$ is the smoothed individual gain of channel m, T is the frame duration, and τ_G is the adaptation time constant, which can be large if the microphone sensitivity doesn't change quickly over time. A prerequisite in order to use this method is a VAD[15] and an isotropic noise field [7]. The alternative to this method is doing the calibration during speech. Again, a VAD is necessary to distinguish between pauses and speech. A prerequisite to use this method is the knowledge of the steering direction ϕ_s, a far-field condition, and a symmetric geometry, which is true in case of a UCA. This method is based on a level interpolation of the microphones' captured energy (see Fig. 2.5b) with a straight line that requires the knowledge of the signal attenuation within the array geometry. It depends on the steering direction ϕ_s and the microphone positions $\boldsymbol{r}_n$. The projection of a microphone onto the line of propagation or DOA[16]-line (see Fig. 2.5a), which passes the origin of the coordinate system, is

$$d_n = r_n \cos(\phi_s - \phi_n).$$

[13] NLMS - Normalized Least Mean Squares
[14] RMS - Root Mean Square
[15] VAD - Voice Activity Detection
[16] DOA - Direction of Arrival



(a) Projection of the mic-positions on the DOA line.

(b) Linear interpolation of the signal levels.

Figure 2.5: Projection of the microphone positions on the DOA line (a) and linear interpolation of the captured energy levels for gain self-calibration (b). The symbol d_i represents the distance between the center of the coordinate system and the microphone projection on the DOA-line.

This distance is also necessary for delay computations (see Section 3.1.2 and Section 3.2.2). According to [8], the level is interpolated as a straight line towards the DOA:

$$\tilde{L}_n(d) = a_{1,n} \cdot d + a_{0,n} = \boldsymbol{a}^T(n)\boldsymbol{d},\ \boldsymbol{a}(n) = \begin{pmatrix} a_1(n) \\ a_2(n) \end{pmatrix},\ \boldsymbol{d} = \begin{pmatrix} d \\ 1 \end{pmatrix}.$$

Solving

$$\min\left(\sum_{m=1}^{N}(\tilde{L}(d_m) - L_m)^2\right)$$

yields the MMSE[17]-solution for the parameters a_1 and a_2 [8], where L_m is the RMS per frame for each channel. The determination of the parameters and levels by considering the interpolation method enables the compensation of the gain-mismatch.

2.4 Short-Time Stationarity of Speech

Because of simplicity, all examples and derivations mentioned before assume monochromatic plane waves, but in real applications broadband signals occur. According to [9] speech exhibits a piecewise short-time stationarity—repeating signal patterns which lead to a quasi-periodic signal for a certain time interval—between $10 - 40$ ms, which is a fundamental assumption for beamforming in frequency domain and in case of block processing.

The maximum array diameter of the UCA in this work is $d = 0.55$ m. A propagating wave passes this distance within

$$\tau = \frac{0.55}{343.2}\frac{m}{m/s} = 0.0016 \text{ s} = 1.6 \text{ ms}.$$

Thus, stationarity of speech can be assumed within the whole array geometry. A sampling

[17] MMSE - Minimum Mean Square Error

frequency of $f_s = 48000$ Hz and a block size of 256 samples results in a 0.0053-seconds time resolution. In this case speech is stationary and quasi-periodic for at least three blocks. Time-alignment in frequency domain and in case of block-processing is efficient only if the signal within a block is (quasi-)periodic.

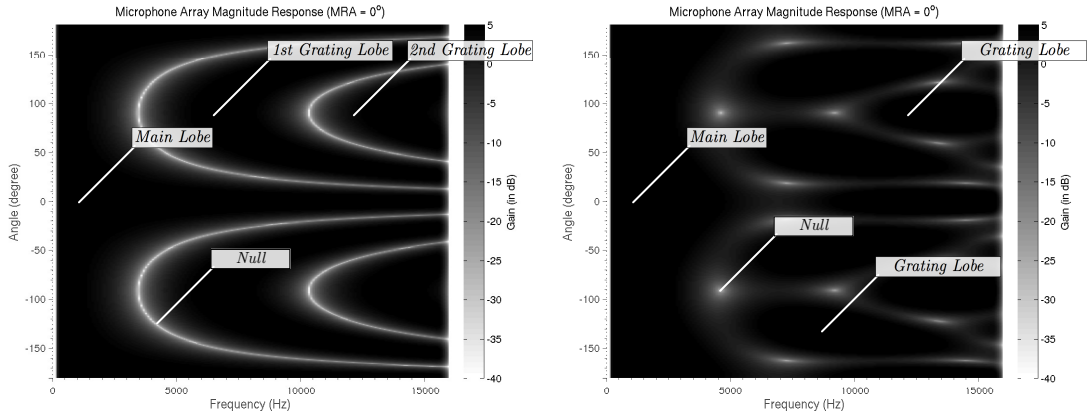
3 Microphone Arrays

3.1 Uniform Linear Array

Microphone arrays do not subject to any restrictions in geometry. There're special types of geometries which are more attractive than randomly positioned microphones because of ordinary and easy-to-implement time-alignment functions. In case of a ULA all microphones are placed on a straight line equidistantly (uniformly). It exhibits a front-back ambiguity because of its linear geometry [7] as shown in Section 3.1.2.

3.1.1 The Array Response

In this work the array response consists of the absolute values of the sums of the steering-direction dependent microphones' sound capture model for each frequency and direction. It describes the directional sensitivity of the microphone array regardless of a beamformer. Fig. 3.1 depicts two typical array responses of a ULA consisting of two (a) and three (b) microphones. The dark area (high gain) on the left part of both plots, separated by bright lines (high attenuation), represents the main lobe; all other separated areas represent grating lobes. The array response exhibits constructive overlapping only for the looking directions $\phi_{l,1} = 0°$ and $\phi_{l,2} = 180°$.



(a) Surface plot of a two-element array.

(b) Surface plot of a three-element array.

Figure 3.1: Both surface plots show the array response of a ULA of two (a) and three (b) microphones with a spacing of $d = 0.05$ m over all frequencies and angles.

3.1.2 The Beam Pattern

Let's consider a ULA which consists of four microphones placed symmetrically around the x-axis of a two-dimensional plane. The simple acoustic pressure field—it consists of a single monochromatic plane wave—is modeled as

$$p_0(\boldsymbol{r}, t) = e^{i(\omega_0 t - \boldsymbol{k}_0^T \boldsymbol{r})}.$$

A ULA without the use of a beamformer exhibits two main lobes at $\phi_{ML1} = 0°$ and $\phi_{ML2} = 180°$. The 1st main lobe arises due to constructive overlapping of the captured waves from the looking direction $\phi_{ML1} = \phi_s = 0°$. The 2nd main lobe appears at $\phi_{ML2} = 180°$ due to the front-back ambiguity (see Fig. 3.2). The captured waves exhibit the same phase, which is not the case for monochromatic plane waves of different frequencies from, e.g., $\phi_s = 1°$ between 100 Hz and 16 000 Hz. A beamformer enables a directional shift of the main lobe in both half planes, whereas the main lobe due to front-back ambiguity is located at

$$\boxed{\phi_{ML2} = 180° - \phi_{ML1}}\ .$$

The introduction of a beamformer (here: a DS-BF aligns the captured waves to obtain constructive overlapping for waves from the desired source or steering direction ϕ_s) requires a virtual reference point, e.g., in the center of the ULA. A close look at Fig. 3.2 reveals that the desired source is located at $\phi_s = 20°$, the competing source—a steadily radiating, undesired source—is positioned at $\phi_c = 160°$. In wavevector notation, there are two monochromatic plane waves arriving from the directions $\boldsymbol{k}_1$ and $\boldsymbol{k}_2$. As shown in Fig. 3.2, the captured waves from both directions experience the same delays due to the steering direction of the DS-BF; thus, it's not possible to determine whether the signal comes from the direction of the first (20°) or the second source (160°), because the steering-direction dependent delays $\tau_{1,1} = \tau_{2,1}$, $\tau_{1,2} = \tau_{2,2}$, and $\tau_{1,3} = \tau_{2,3}$ are identical in both cases. Each delay depends on the distance between the array reference point and a microphone $\boldsymbol{r}_n$, the steering direction ϕ_s, and the speed of sound c.

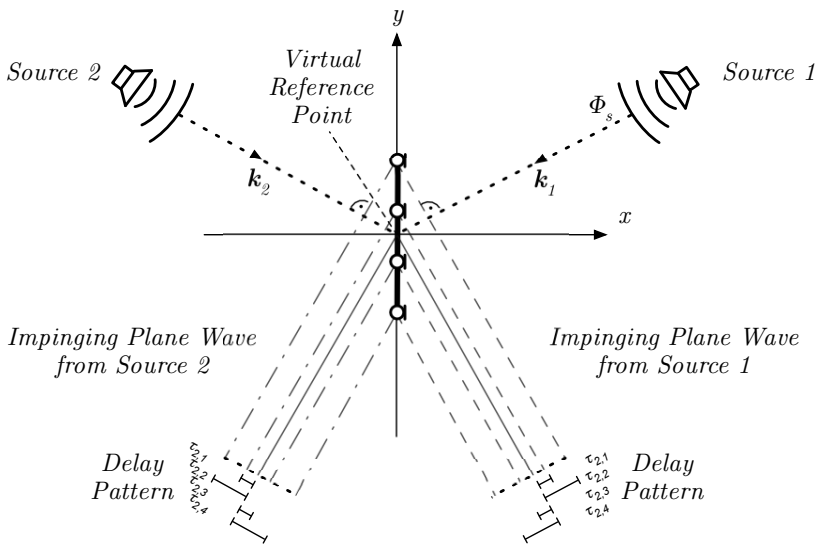


Figure 3.2: A ULA, which consists of four microphones, receives signals from two different directions: 20° *and* 160°. *Although both sources are positioned at different places, the delays* $\tau_{1,1} = \tau_{2,1}$, $\tau_{1,2} = \tau_{2,2}$, *and* $\tau_{1,3} = \tau_{2,3}$ *are identical in both cases. The delay pattern unveils information about the necessary delays for each microphone to obtain constructive overlapping for a certain direction.*

In general, the array reference point is in the center of the coordinate system and the ULA. The delay is calculated as follows [4][18]: inserting $\boldsymbol{k}_0^T \boldsymbol{r}_n$ into the beamformer's output (2.7) results in

$$y(t) = \frac{1}{N}\sum_{n=1}^{N} e^{i(\omega_0[t-\tau_n]-\boldsymbol{k}_0^T\boldsymbol{r}_n)} = e^{i\omega_0 t}\cdot\frac{1}{N}\sum_{n=1}^{N} e^{-i(\omega_0\tau_n+\boldsymbol{k}_0^T\boldsymbol{r}_n)}$$

$$y(t) = e^{i\omega_0 t}\cdot\frac{1}{N}\sum_{n=1}^{N}\underbrace{e^{-i\omega_0\tau_n}}_{B_n(\omega_0)}\underbrace{e^{i\frac{\omega_0}{c}r_n\cos(\phi_s-\phi_n)}}_{D_n(\omega_0,\phi_s)}$$

in time domain—$B_n(\omega_0)$ is a beamformer kernel—and

$$Y(\omega,\varphi,\phi_s) = 2\pi\delta(\omega-\omega_0)\frac{1}{N}\sum_{n=1}^{N} e^{-i\omega_0\tau_n}e^{i\frac{\omega_0}{c}r_n\cos(\phi_s-\phi_n)}$$

in frequency domain, and the beam pattern—it is the sum of multiplied array response and beamformer kernels—for the steering direction ϕ_s is

$$H(\omega,\varphi,\phi_s) = \frac{1}{N}\sum_{n=1}^{N} e^{-i\frac{\omega}{c}r_n\cos(\varphi-\phi_n)}\cdot e^{i\frac{\omega}{c}r_n\cos(\phi_s-\phi_n)} = \frac{1}{N}\sum_{n=1}^{N} B_n(\omega)\cdot D_n(\omega),$$

$$\boxed{H(\omega,\varphi,\phi_s) = \frac{1}{N}\sum_{n=1}^{N} e^{i\frac{\omega}{c}r_n(\cos(\phi_s-\phi_n)-\cos(\varphi-\phi_n))}}, \tag{3.1}$$

for general ω. That implies that the frequency-independent delay is

$$\tau_n = \frac{r_n\cdot\cos(\varphi-\phi_n)}{c}. \tag{3.2}$$

where c is the sound velocity, r_n is the distance between the virtual reference point and the microphones, ϕ_n is the microphone angle, ϕ_s is the steering direction and the desired source angle, and φ is the delay compensating angle, which should be the source angle for a perfect capture of signals from this direction. Fig. 3.3a depicts a 4-element ULA with steering direction $\phi_s = 25°$. The virtual mapping line—it passes the virtual reference point and is perpendicular to ϕ_s—splits the 2-dimensional plane into two half planes, whereas the waves captured by the microphones in the right plane experience a delay in time (i.e. a causal system behaviour), and all others experience an advance in time (i.e. a non-causal system behaviour) due to the use of a beamformer. Waves from the competing source direction ϕ_c do not experience any delays that lead to constructive overlapping; but there are constructive interferences for a countable number of signal components of a broadband signal due to a perfect match of the captured signal-component's phase, their wavelengths, and the beamformer's delays. The delays are equivalent to 2π-phase-rotations or inter multiples of it. Fig. 3.3b illustrates a virtual mapping line for waves from $\phi_c = 0°$. In case of a continuous array it unveils information about the alignment for each sensor. Constructive overlapping occurs if the virtual mapping line of ϕ_c is parallel to the virtual mapping line of ϕ_s. Because of a lack of derivations of the implementation mentioned above, the closed-form solution of the beam pattern is derived in this work. It is crucial to determine an analytical description of the distance between the virtual reference point and each

[18] Note: In comparison to [4] the whole coordinate system is shifted $-90°$.

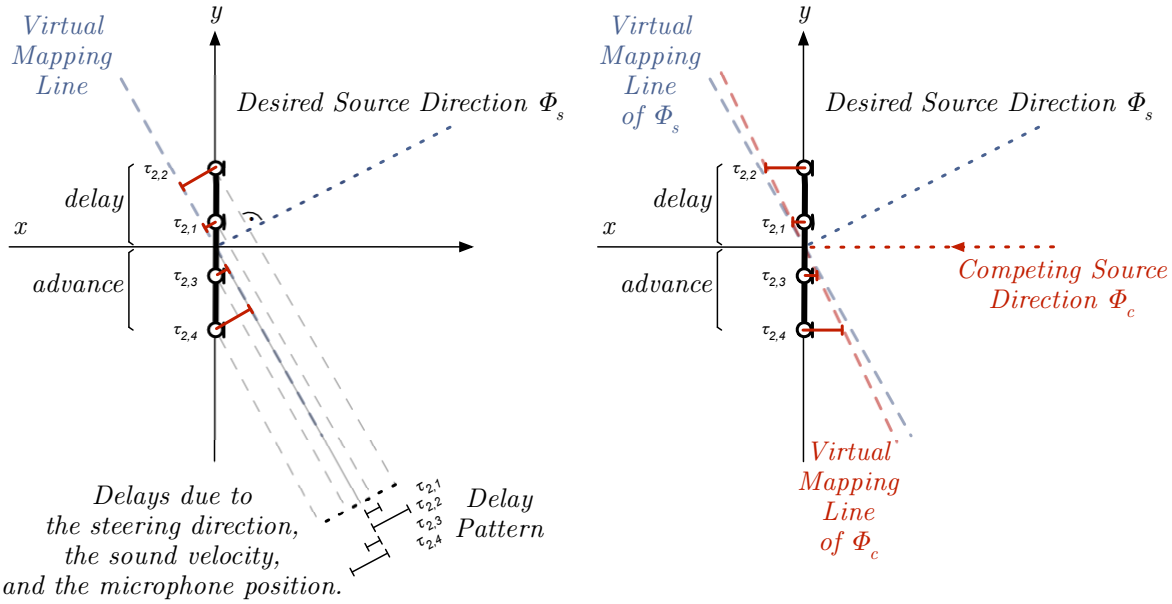


Figure 3.3: The left figure shows a delay pattern for signals from ϕ_s captured by all microphones. The right figure shows the virtual mapping line of both sources, the desired and competing source.

microphone.

$$r_n = d\left(\frac{N-1}{2} - n\right),\ n = \{0, 1, 2, ..., N/2-1\},\ N = 2k,\ k \in \mathbb{N} \tag{3.3}$$

Rewriting the beam pattern leads to

$$H(\omega, \varphi, \phi_s) = \frac{1}{N}\sum_{n=0}^{N/2-1} e^{i\frac{\omega}{c}d\left(\frac{N-1}{2}-n\right)(\cos(\phi_s-\phi_{pos})-\cos(\varphi-\phi_{pos}))} + \frac{1}{N}\sum_{n=0}^{N/2-1} e^{-i\frac{\omega}{c}d\left(\frac{N-1}{2}-n\right)(\cos(\phi_s-\phi_{neg})-\cos(\varphi-\phi_{neg}))},$$

where $\phi_{pos} = 90°$ and $\phi_{neg} = -90°$ are the microphone angles. This yields

$$H(\omega, \varphi, \phi_s) = \frac{1}{N}\sum_{n=0}^{N/2-1} e^{i\frac{\omega}{c}d\left(\frac{N-1}{2}-n\right)(\sin(\phi_s)-\sin(\varphi))} + \frac{1}{N}\sum_{n=0}^{N/2-1} e^{-i\frac{\omega}{c}d\left(\frac{N-1}{2}-n\right)(\sin(\phi_s)-\sin(\varphi))}.$$

In comparison to (3.1) the beam pattern equation is now split into two parts, because (3.3) has to be positive. If index n goes until $N-1$, r_n becomes negative for $n > N/2-1$, which compensates the sign generated by the expression $\cos(\phi_s - \phi_{neg}) - \cos(\varphi - \phi_{neg})$. Further

simplifications by considering substitutions yield

$$H(\eta) = \frac{1}{N}\left(e^{i\eta\frac{N-1}{2}} \cdot \sum_{n=0}^{N/2-1} e^{-i\eta n} + e^{-i\eta\frac{N-1}{2}} \cdot \sum_{n=0}^{N/2-1} e^{i\eta n} \right), \tag{3.4}$$

where

$$\eta = \frac{\omega}{c} d \cdot (\sin(\phi_s) - sin(\varphi)) .$$

Equation (3.4) can be rewritten by considering

$$\sum_{n=0}^{N/2-1} x^n = \frac{1 - x^{\frac{N}{2}}}{1-x},$$

which yields a closed-form expression that describes the beam pattern in terms of a Dirichlet kernel [10]

$$\begin{aligned} H(\eta) &= \frac{1}{N}\left(e^{i\eta\frac{N-1}{2}} \left[\frac{1 - e^{-i\eta\frac{N}{2}}}{1 - e^{-i\eta}}\right] + e^{-i\eta\frac{N-1}{2}} \left[\frac{1 - e^{i\eta\frac{N}{2}}}{1 - e^{i\eta}}\right] \right) \\ &= \frac{1}{N}\left(\frac{e^{i\frac{N}{2}\eta} - 1}{e^{i\frac{\eta}{2}} - e^{-i\frac{\eta}{2}}} + \frac{e^{-i\frac{N}{2}\eta} - 1}{e^{-i\frac{\eta}{2}} - e^{i\frac{\eta}{2}}} \right) = \frac{1}{N}\left(\frac{e^{i\frac{N}{2}\eta} - 1}{e^{i\frac{\eta}{2}} - e^{-i\frac{\eta}{2}}} - \frac{e^{-i\frac{N}{2}\eta} - 1}{e^{i\frac{\eta}{2}} - e^{-i\frac{\eta}{2}}} \right) \\ &= \frac{1}{N}\left(\frac{e^{i\frac{N}{2}\eta} - 1 - e^{-i\frac{N}{2}\eta} + 1}{e^{i\frac{\eta}{2}} - e^{-i\frac{\eta}{2}}} \right) = \frac{1}{N}\frac{2i}{2i}\left(\frac{e^{i\frac{N}{2}\eta} - e^{-i\frac{N}{2}\eta}}{e^{i\frac{\eta}{2}} - e^{-i\frac{\eta}{2}}} \right) \end{aligned}$$

$$\boxed{H(\eta) = \frac{1}{N}\frac{\sin\left(\frac{N}{2}\eta\right)}{\sin\left(\frac{\eta}{2}\right)} = \frac{1}{N}\frac{\sin\left(\frac{N\cdot\frac{2\pi}{\lambda}d\cdot(\sin(\phi_s)-sin(\varphi))}{2}\right)}{\sin\left(\frac{\frac{2\pi}{\lambda}d\cdot(\sin(\phi_s)-sin(\varphi))}{2}\right)}} \tag{3.5}$$

for an even and odd number of microphones. In case of an odd number of microphones, the derivation is different to the previous derivation, i.e.

$$r_n = d\left(\frac{N-1}{2} - n\right),\ n = \{0, 1, 2, ..., (N-1)/2\},\ N = 2k+1,\ k \in \mathbb{N}$$

and

$$\begin{aligned} H(\omega, \varphi, \phi_s) = \frac{1}{N} \sum_{n=0}^{(N-1)/2} e^{i\frac{\omega}{c}d\left(\frac{N-1}{2}-n\right)(\cos(\phi_s-\phi_{pos})-\cos(\varphi-\phi_{pos}))} - \frac{1}{N} \\ + \frac{1}{N} \sum_{n=0}^{(N-1)/2} e^{-i\frac{\omega}{c}d\left(\frac{N-1}{2}-n\right)(\cos(\phi_s-\phi_{neg})-\cos(\varphi-\phi_{neg}))}, \end{aligned}$$

which yields

$$H(\eta) = \frac{1}{N}\left((-1) + e^{i\eta\frac{N-1}{2}} \left[\frac{1 - e^{-i\eta\frac{N+1}{2}}}{1 - e^{-i\eta}}\right] + e^{-i\eta\frac{N-1}{2}} \left[\frac{1 - e^{i\eta\frac{N+1}{2}}}{1 - e^{i\eta}}\right] \right) = ... = \frac{1}{N}\frac{\sin\left(\frac{N}{2}\eta\right)}{\sin\left(\frac{\eta}{2}\right)},$$

otherwise the captured waves from the microphone in the center of the coordinate system are

considered twice. The first null in the beam pattern can be calculated by setting the argument of $\sin(\frac{N}{2}\eta)$ to $\pm\pi$. If the arguments of both sine-functions are identical, a grating lobe—a lobe with the same maximum gain or a higher gain as the main lobe—occurs in the beam pattern. If the argument of the sine-function in the numerator is $\pm\frac{\pi}{2}$ and not equal to the argument of the sine-function in the denominator, a side lobe—a lobe with a gain smaller than the gain of the main lobe—occurs.

3.1.3 Spatial Aliasing and Grating Lobes due to Phase Ambiguities

According to [7] spatial aliasing arises when the resulting phase shift in a single microphone-pair–combination is identical for two or more different directions for a given frequency. For instance, a monochromatic plane wave with frequency f from a specific direction passes two microphones in such a way that the captured waves exhibit the same phase, so that $s_1(t) = s_2(t) = s(t)e^{\pm ia}$, which leads to constructive overlapping after summarizing both signals. In case of a ULA, an increase in the number of microphones without changing the microphone spacing decreases the spatial aliasing frequency f_{sa}, because the distance between the microphones at both ends of the ULA increases which shifts the 1st side lobe to lower frequencies. By contrast, a decrease in the microphone spacing without changing the number of microphones increases the spatial resolution, and, thus, shifts f_{sa} to a higher frequency. A grating lobe arises due to the identical phase shift in all microphones. The gain of these grating lobes matches the gain or is higher than the gain of the main lobe. These lobes cause the microphone array to capture signal components from different directions without attenuation. Fig. 3.4 depicts the array response of a 2-element ULA for different frequencies. For $f = 6864$ Hz the array response exhibits a grating lobe, where the attenuation of the lobe at 0° is the same as the attenuation of the grating lobe at 90°: 0 dB. Let's consider a ULA consisting of two omnidirectional microphones without the use of a beamformer. The microphones are placed symmetrically on the y-axis, and the distance between

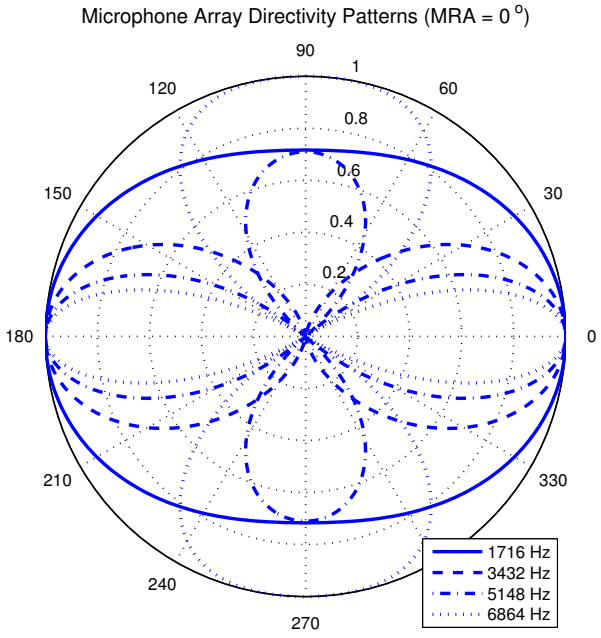


Figure 3.4: Array response for different frequencies of a ULA, which consists of two microphones, and which exhibits a looking direction of 0°.

both microphones is d. The sound capture model of a microphone is described as follows:

$$D_n(\omega, \phi_s) = e^{i\frac{\omega}{c}d \cdot cos(\phi_s - \phi_n)}, \tag{3.6}$$

where $n \in \{1, 2\}$ is the microphone index, f is the observed frequency, ϕ_n is microphone angle, and ϕ_s is the steering direction. This model describes the captured monochromatic plane waves by multiplying the model's frequency response with the Fourier transform of $r(t)$, i.e. $R_1(\omega) = R(\omega)D_1(\omega)$, where $r(t)$ is the source signal. Now, let's have a focus on the exponent of the capturing model, which contains information about the phase of the received waves.

$$\frac{\omega}{c}d \cdot cos(\phi_s - \phi_n) = \frac{2\pi}{\lambda}d \cdot cos(\phi_s - \phi_n) \tag{3.7}$$

Equation (3.7) considers the distance between both microphones, which is an essential quantity for determining the grating lobe frequency f_{gl}, which is depicted in Fig. 3.5.

Spatial aliasing and grating lobes occur because the waves captured with both microphones exhibit the same frequency and the same phase which results in constructive overlapping after summarizing both waves. The frequency of the grating lobe can be calculated as follows:

$$d \cdot \cos(\phi_s - 90°) \stackrel{!}{=} m \cdot \lambda_{gl} = \frac{c \cdot m}{f_{gl}}, \tag{3.8}$$

where m describes the number of the grating lobe and $m \in \mathbb{N}$. Rewriting (3.8) yields

$$f_{gl} = \frac{c \cdot m}{d \cdot \cos(\phi_s - 90°)}$$

and for general microphone angles (here: $0 \leq \phi_n \leq \pi$)

$$\boxed{f_{gl} = \frac{c \cdot m}{d \cdot \cos(\phi_s - \phi_n)}}\,. \tag{3.9}$$

The result changes slightly if we consider a DS-BF. The system response for the microphone with index n is

$$H_n(\omega, \varphi) = e^{i\frac{\omega}{c}r_n(\cos(\phi_s - \phi_n) - \cos(\varphi - \phi_n))},$$

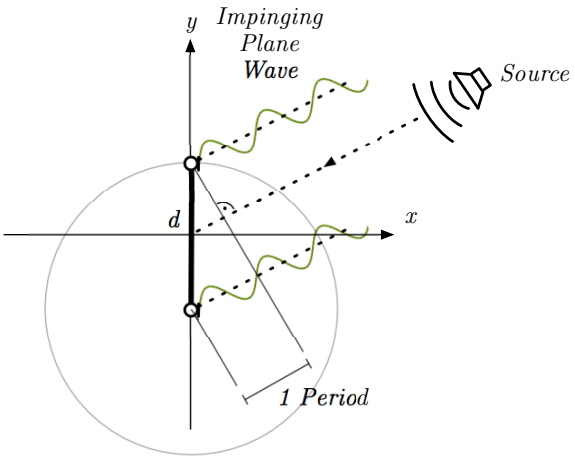


Figure 3.5: In this figure a grating lobe occurs for a certain frequency and its integer multiple by considering a two-element microphone array with distance d between both microphones.

which consists of the sound capture model $D_n(\omega, \phi_s) = e^{i\frac{\omega}{c} r_n \cos(\phi_s - \phi_n)}$ and the beamformer response $B_n(\omega, \varphi) = e^{-i\frac{\omega}{c} r_n \cos(\varphi - \phi_n)}$. Considering both responses yields

$$\boxed{f_{gl} = \frac{c \cdot m}{d \cdot [\cos(\phi_s - \phi_n) - \cos(\varphi - \phi_n)]}} \,. \tag{3.10}$$

Both equations, (3.9) and (3.10), are independent of the number of microphones in case of equidistantly distributed microphones, but they depend on the microphone spacing. Thus, using additional microphones with the same sensor interval does not change the frequency of the maximum point of the grating lobes, but a change in microphone spacing will do so.

3.1.4 Characteristics of the Uniform Linear Array

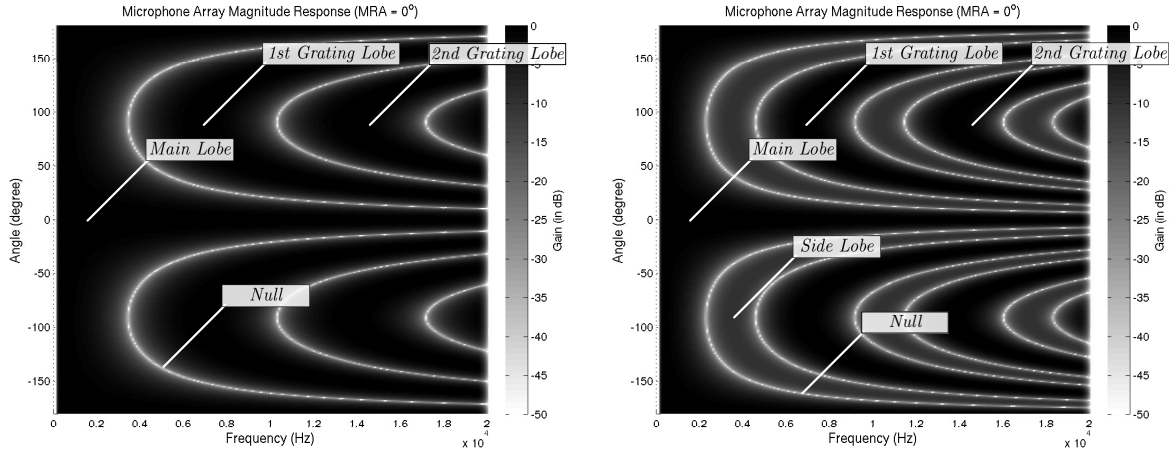
Fig. 3.6 to Fig. 3.8 depict two surface plots, two one-dimensional plots and two polar plots of an array response of a ULA consisting of two (a) and three (b) microphones with a microphone spacing of $d = 0.05$m and a looking direction of $\phi_l = 0°$ (see Fig. 2.4) regardless of a beamformer but with consideration of a microphone-number dependent scaling factor.

Let's start with both surface plots (a) and (b) of Fig. 3.6. Their x-axes represent the frequencies ranging from 0 Hz to 20000 Hz, the y-axes represent the azimuth etween $-180° \leq \varphi \leq +180°$, and the color bars provide information about the attenuation of frequency components of signals from different directions. The left array response (a) exhibits a single main lobe and no side lobes, all other lobes are grating lobes (see comments in plots). A thin line exhibiting a high attenuation—it is illustrated as a bright line—separates the lobes. The right array response (b) shows a main lobe, three side lobes—they exhibit an increase in brightness and they never reach a gain equal to 0 dB—and three grating lobes. The brighter the area the higher the attenuation. The lowest attenuation is 0 dB, the highest possible attenuation equals $-\infty$ dB. Plots (a) and (b) of Fig. 3.7 show the array response for certain frequencies (4000 Hz, 5000 Hz, 6000 Hz, and 7000 Hz) over all angles. Plot (a) shows a gain smaller than one at 4000 Hz and 90°, which seems to be a side lobe at the first sight. Indeed, it's a grating lobe because the gain reaches 0 dB at higher frequencies and it behaves as a strictly monotonic increasing function until 0 dB. In (b) one can see a side lobe developing from another one at 4000 Hz and 90°. The minimum attenuation of both lobes is -10 dB, i.e. it is less than 0 dB, and thus it cannot be a grating lobe. Both polar plots (a) and (b) of Fig. 3.8 provide the same information as (a) and (b) of the previous figure except that the attenuation is plotted linearly and not logarithmically.

Fig. 3.9 and Fig. 3.10 depict the same type of plots of a ULA consisting of four (a, c, e) and five (b, d, f) microphones, but with a sensor interval of $d = 0.05$m and $d = 0.025$m (in this order) and a looking direction of $\phi_l = 0°$. It's interesting to see in Fig. 3.9 or Fig. 3.10 that using four and five microphones results in a array response with two and three side lobes, a thinner main lobe (a decrease in angles), and smaller grating lobes (a decrease of the frequency range) in comparison to Fig. 3.6 to Fig. 3.8. Thus, an increase in the number of microphones without a change in the microphone spacing leads to

- an increased number of side lobes,
- an increased attenuation of the side lobes,
- a decrease in the main, side, and grating lobe width,
- an increase in frequency f_{gl} of the grating lobe maximum (0 dB), and
- an increased spatial aliasing frequency f_{sa}.

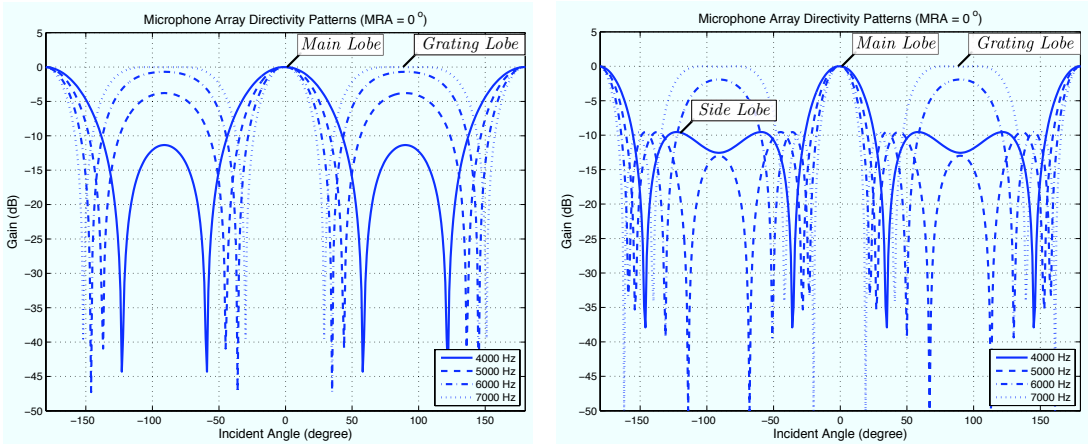
A decrease in the the microphone spacing without a change of the number of microphones yields



(a) Surface plot of a two-element array.

(b) Surface plot of a three-element array.

Figure 3.6: The surface plots show the array response for all frequencies and angles. Computations are based on a ULA consisting of two (a) and three (b) microphones with a microphone spacing of $d = 0.05m$.



(a) One-dimensional plot of a two-element array.

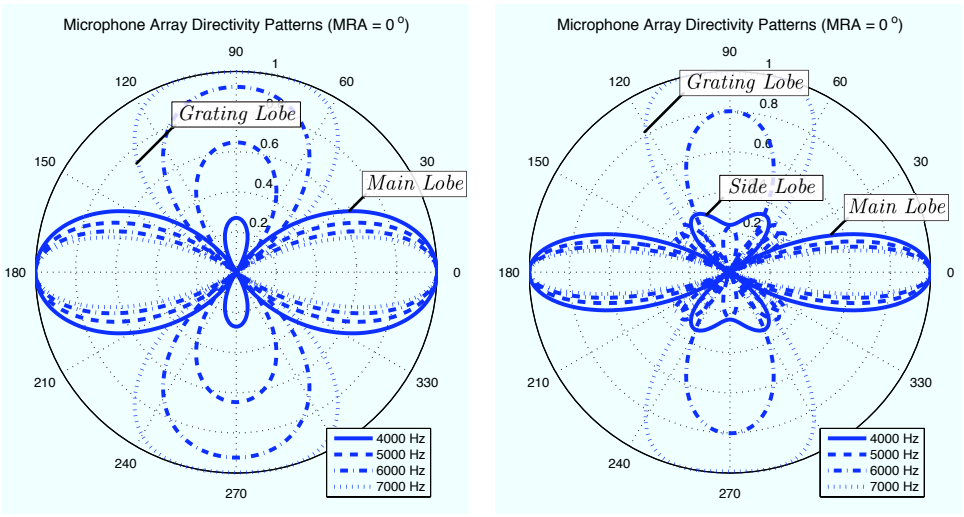
(b) One-dimensional plot of a three-element array.

Figure 3.7: The one-dimensional plots show the array response for all angles but a certain number of frequencies. Computations are based on a ULA consisting of two (a) and three (b) microphones with a microphone spacing of $d = 0.05m$.

- an increase in the main, side, and grating lobe width,
- an increase in frequency f_{gl} of the grating lobe maximum (0 dB), and
- an increased spatial aliasing frequency f_{sa}.

In general, the minimum attenuation of, e.g., the first side lobe in case of an array consisting of four microphones is the same for both sensor intervals: $d = 0.05$ m and $d = 0.025$ m.

In the examples mentioned above the ULA is used in broadside mode. In the example shown in Fig. 3.11 the microphone array is used as an end-fire mode which exhibits a steering direction along the array axis of $\phi_s = \phi_n = 90°$, and which requires a beamformer (here: DS-BF). The use of a ULA in end-fire mode yields



(a) Polar plot of a two-element array.

(b) Polar plot of a three-element array.

Figure 3.8: The polar plots show the directivity pattern. Computations are based on a ULA consisting of two (a) and three (b) microphones with a microphone spacing of $d = 0.05m$.

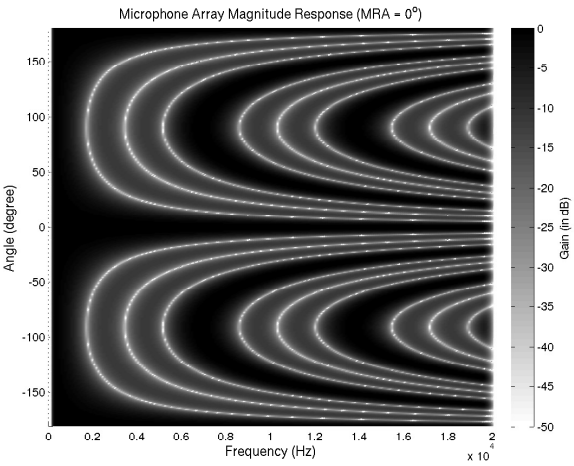
- an elimination of the front-back ambiguity,
- an increase in the main and side lobe width,
- a decrease in frequency f_{gl} of the grating lobe maximum, and
- an increase of the spatial aliasing frequency f_{sa}.

The increased width of the main lobe is due to the overlapping of the main lobe in steering direction ϕ_s and the lobe caused by front-back ambiguity.

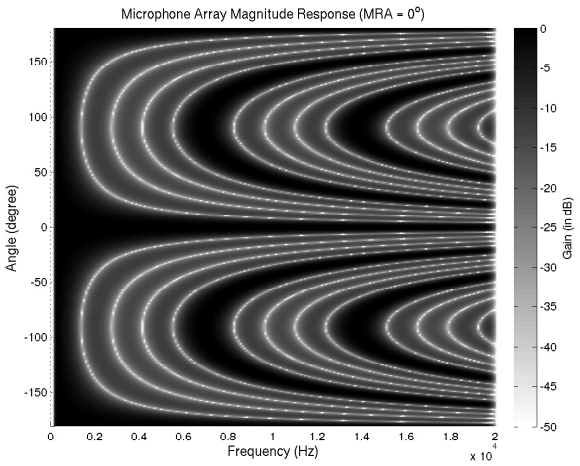
Based on all observations, the number of side lobes between the main lobe and the first grating lobe for linear arrays can be defined as

$$\boxed{SL_{num} = N - 2}\,, \tag{3.11}$$

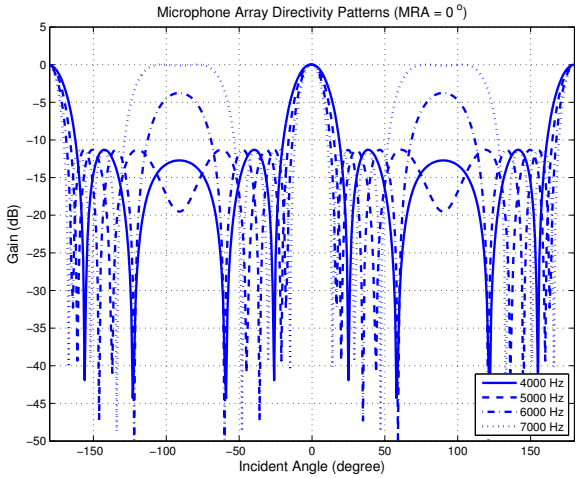
where N is the number of microphones. Equation (3.11) is true for different equidistant sensor intervals because perfect overlapping has to occur in all microphone-pair–combinations which depends on the number of microphones.



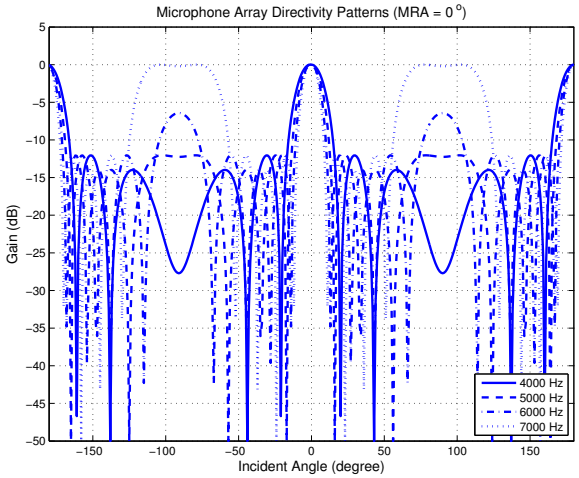
(a) Surface plot of a four-element array.



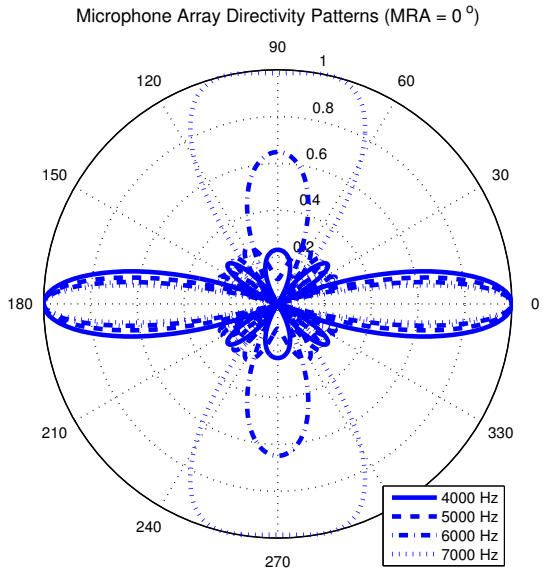
(b) Surface plot of a five-element array.



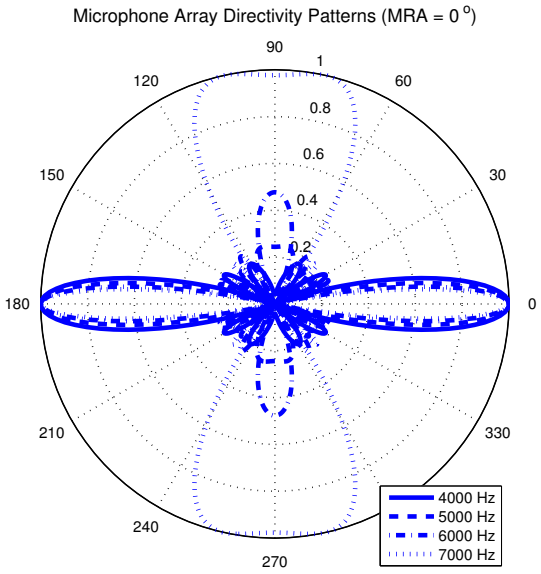
(c) One-dimensional plot of a four-element array.



(d) One-dimensional plot of a five-element array.

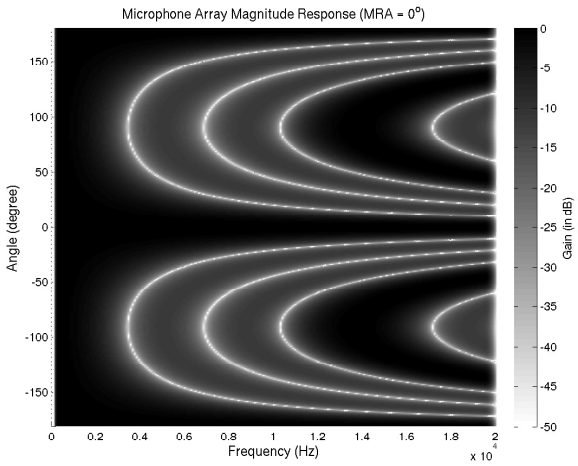


(e) Polar plot of a four-element array.

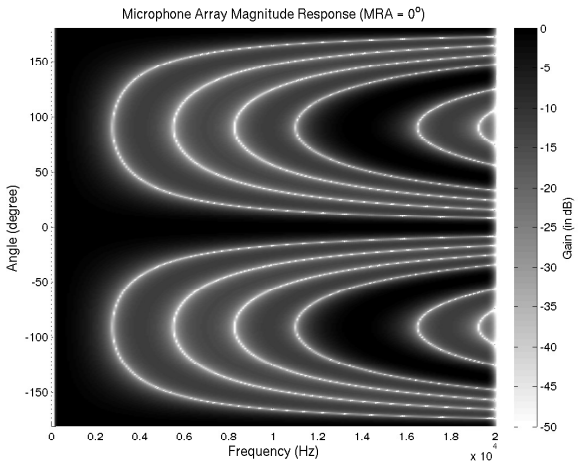


(f) Polar plot of a five-element array.

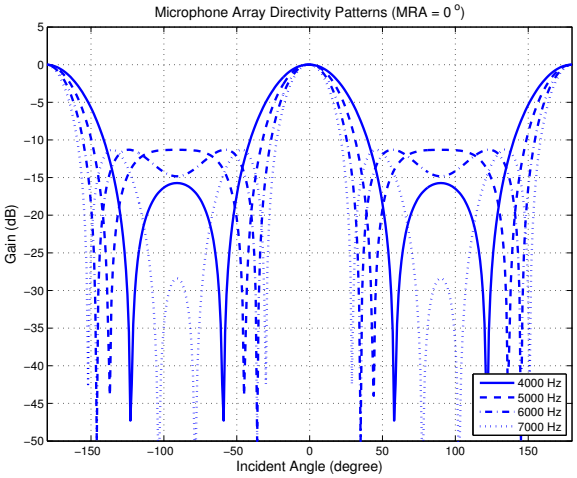
Figure 3.9: The surface plots (a-b) show the array response for all frequencies and angles. The one-dimensional plots (c-d) show the array response for the given frequencies $f = \{4000, 5000, 6000, 7000\}$ *Hz, and so are the polar plots (e-f). Computations are based on a ULA consisting of four (a,c,e) and five (b,d,f) microphones. The distance between all microphones is* $d = 0.05$ *m.*



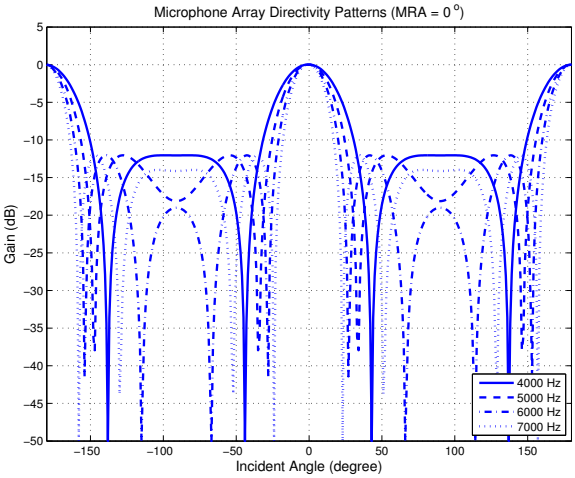
(a) Surface plot of a four-element array.



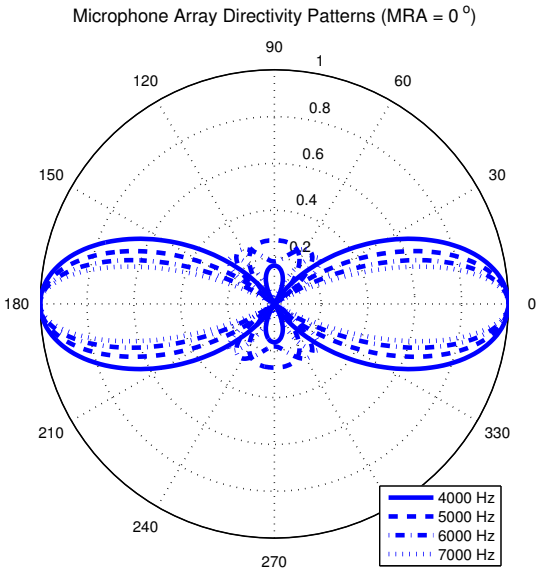
(b) Surface plot of a five-element array.



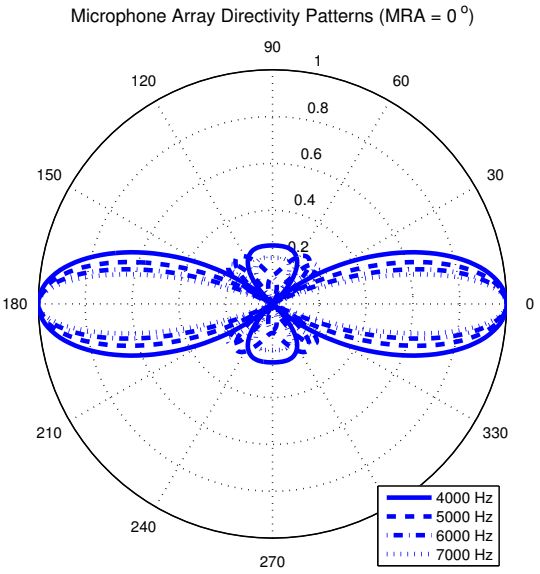
(c) One-dimensional plot of a four-element array.



(d) One-dimensional plot of a five-element array.

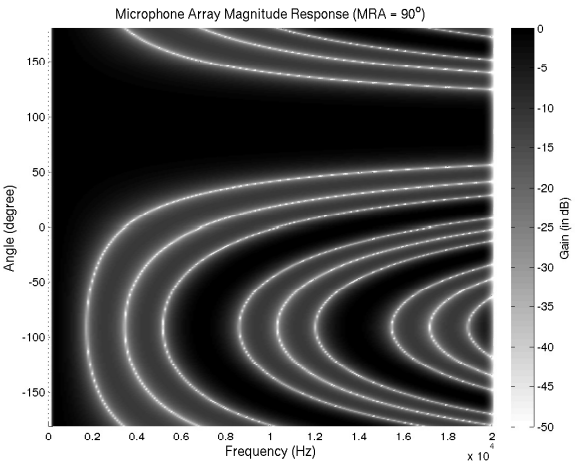


(e) Polar plot of a four-element array.

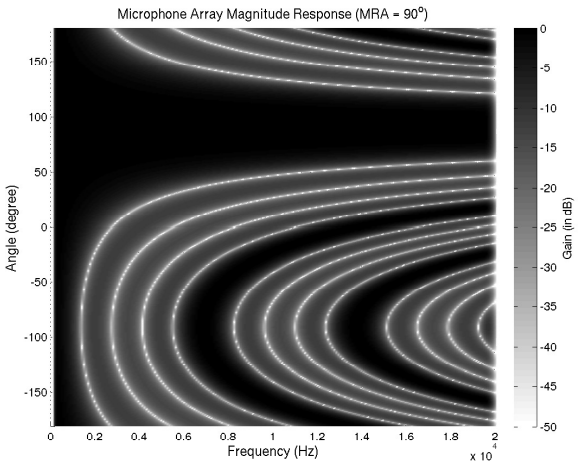


(f) Polar plot of a five-element array.

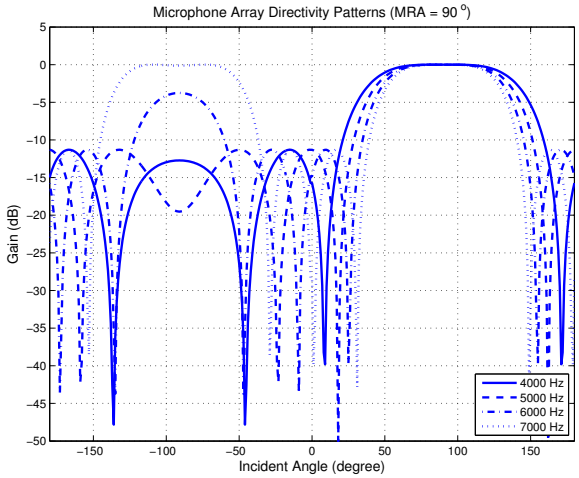
Figure 3.10: The surface plots (a-b) show the array response for all frequencies and angles. The one-dimensional plots (c-d) show the array response for the given frequencies $f = \{4000, 5000, 6000, 7000\}$ *Hz, and so are the polar plots (e-f). Computations are based on a ULA consisting of four (a,c,e) and five (b,d,f) microphones. The distance between all microphones is* $d = 0.025$ *m.*



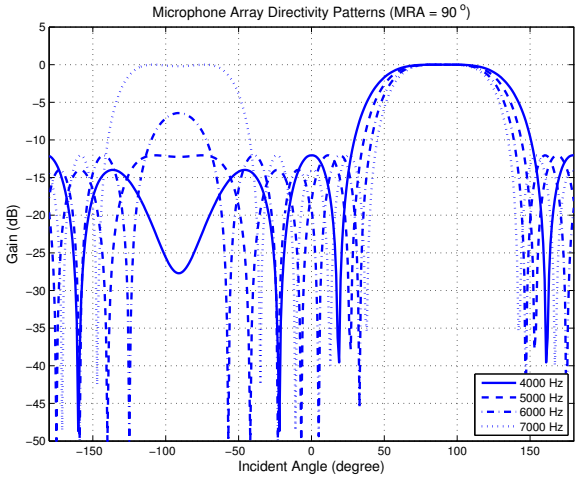
(a) Surface plot of a four-element array.



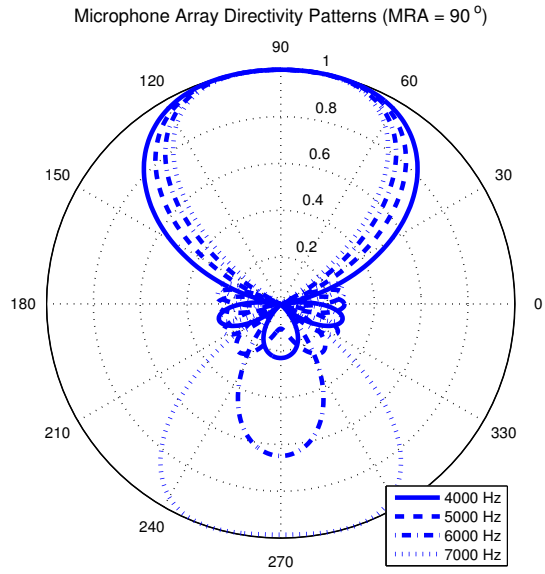
(b) Surface plot of a five-element array.



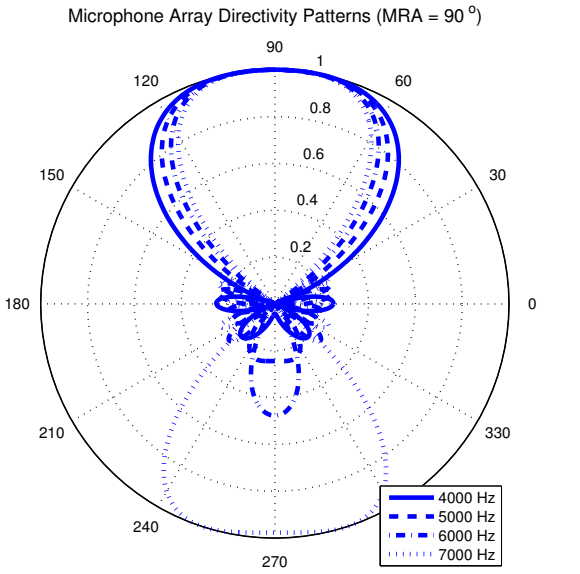
(c) One-dimensional plot of a four-element array.



(d) One-dimensional plot of a five-element array.



(e) Polar plot of a four-element array.



(f) Polar plot of a five-element array.

Figure 3.11: This figure shows the beam pattern for all angles and (all) frequencies. Computations are based on a ULA consisting of four (a,c,e) and five (b,d,f) microphones. The distance between all microphones is $d = 0.025$ m and—in comparison to the previous figures—the steering direction is set to $\phi_s = 90°$ which implies the use of a beamformer (here: DS-BF).

3.2 Uniform Circular Array

The use of a ULA in source localization causes front-back ambiguity. Building up such an array in the center of a room enables source localization in a single half plane only, because the algorithms used for source localization are not able to determine whether the sound source is positioned in front of or behind the ULA (see Fig. 3.2), whereas a planar array, e.g., a UCA, enables source localization in both half planes.

3.2.1 The Array Response

The superposition of the 'complex' array responses of two ULAs with $\phi_{l1} = 0°$ and $\phi_{l2} = 90°$ results in the array response of a 4-element UCA shown in Fig. 3.12a. The superposition of the 'complex' array responses of four ULAs with $\phi_{l1} = 0°$, $\phi_{l2} = 45°$, $\phi_{l3} = 90°$, and $\phi_{l4} = 135°$ results in the array response of a 4-element UCA shown in Fig. 3.12b. Thus, a decomposition of a UCA into microphone pairs yields a certain number of 2-element ULAs with different looking directions $\phi_{l,i}$.

3.2.2 The Beam Pattern

Closed-form solutions for beam patterns of planar arrays are scarce, too complex, or simply non-existent. Therefore, this work is primarily concerned with the beam pattern of the UCA without deriving a closed-form solution.

The UCA exhibits a constant radius between the center of the coordinate system and the microphones, and all microphones feature different but equidistantly distributed angles. Since the equation, which describes the beam pattern of the ULA, is derived for general radii and microphone positions, it can be used to describe the beam pattern of the UCA too:

$$\boxed{H(\omega, \varphi, \phi_s) = \frac{1}{N} \sum_{n=1}^{N} e^{i\frac{\omega}{c} r_n (\cos(\phi_s - \phi_n) - \cos(\varphi - \phi_n))}}, \tag{3.12}$$

where

$$r_n = r$$

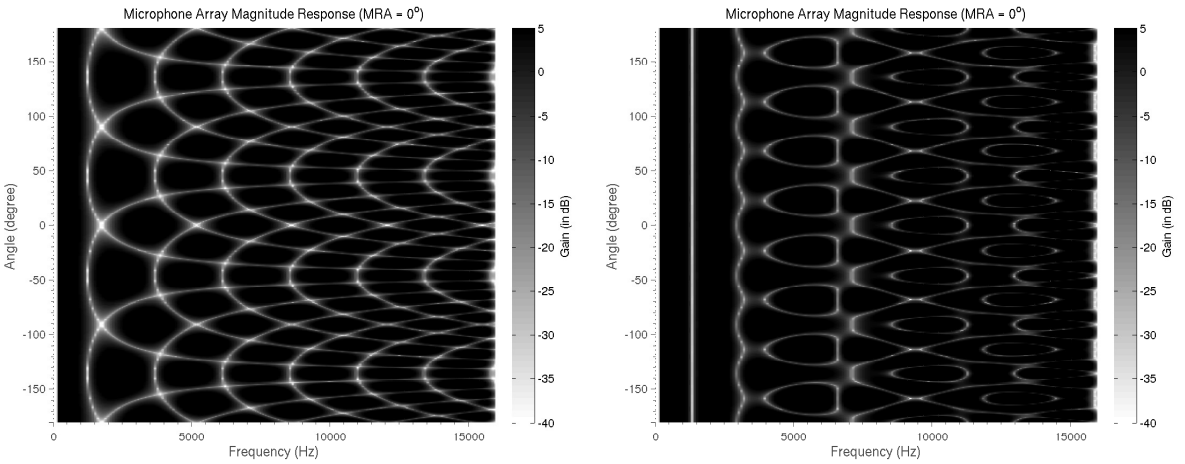
and

$$\phi_n = \frac{2\pi n}{N} \cdot \frac{180}{\pi},\ n = \{0, 1, 2, ..., N-1\},\ N \in \mathbb{N}$$

or

$$\phi_n = \frac{2\pi (n-1)}{2N} \cdot \frac{180}{\pi},\ n = \{1, 2, ..., 2N\},\ N \in \mathbb{N}.$$

The virtual mapping line splits the 2-dimensional plane into two half planes, whereas the monochromatic plane waves captured by the microphones in the source-including half plane experience a delay in time (i.e. a causal system behaviour), and all others experience an advance in time (i.e. a non-causal system behaviour) due to the use of a beamformer. A close look at Fig. 3.13 and a focus on the half-plane closest to the desired source with steering direction ϕ_s reveals that all microphones within this area sample the sound field on certain positions. Summarizing all captured signals without manipulating the phase doesn't lead to constructive



(a) Surface plot of a four-element array.

(b) Surface plot of a eight-element array.

Figure 3.12: The surface plots show the array response for all frequencies and angles. Computations are based on a UCA consisting of four (a) and eight (b) microphones with a diameter of $d = 0.20m$.

overlapping. The beamformer has to shift the signals in time, so that all signal components from the steering direction ϕ_s overlap constructively. The time-shift depends on the speed of sound c, the constant radius r, the microphone positions $\boldsymbol{r}_n$, and the steering direction ϕ_s. The frequency independent delay is modeled as

$$\tau_n = \frac{r}{c}\cos(\varphi - \phi_n) \ .$$

All microphones in the remaining half-plane capture the impinging waves, and the beamformer advances all waves from the direction ϕ_s in time. Advancing indicates non-causality, which actually evokes no problems because of block-processing and short-time stationarity of speech.

Due to delaying in the right half-plane and advancing in the left one, signal components from direction ϕ_s overlap constructively, and signal components from direction ϕ_c are advanced in the left half-plane and delayed in the right one. Fig. 3.14 show the corresponding delay pattern for monochromatic plane waves from the steering direction ϕ_s and from the competing source direction ϕ_c, which is shifted 180 degrees compared to the steering direction ϕ_s.

The beamformer shifts the captured waves from ϕ_s on the virtual mapping line, where they exhibit the same phase and, thus, overlap constructively. Waves from other directions experience the same shifts; but they are mapped on an ellipsoid instead of a straight line which is necessary for constructive overlapping.

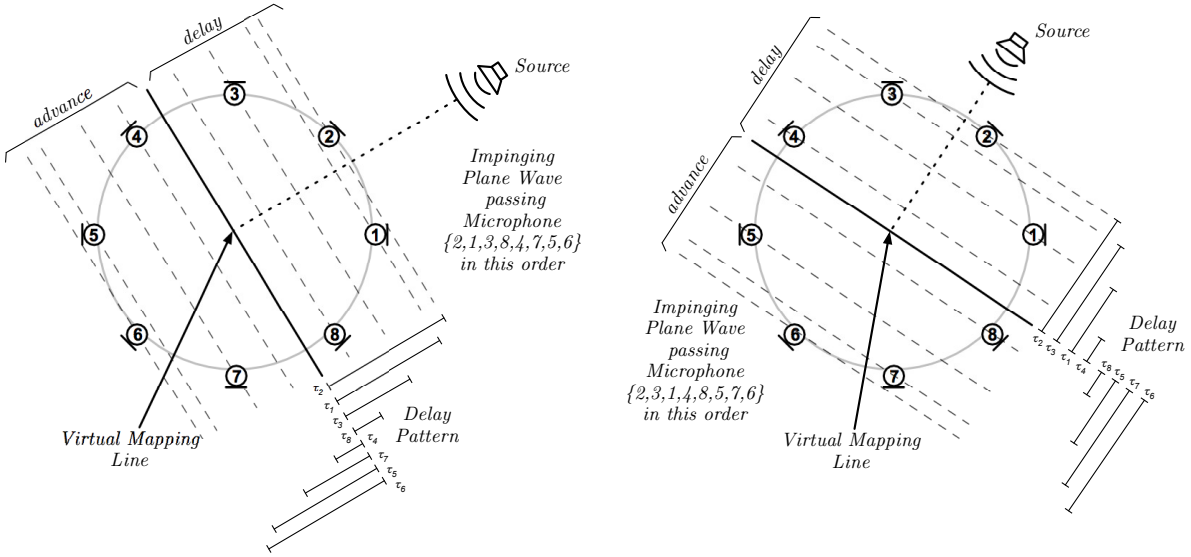


Figure 3.13: Structure of a UCA with radius r and 8 microphones. The bold diagonal line is the virtual mapping line, which is the reference line for delay computations.

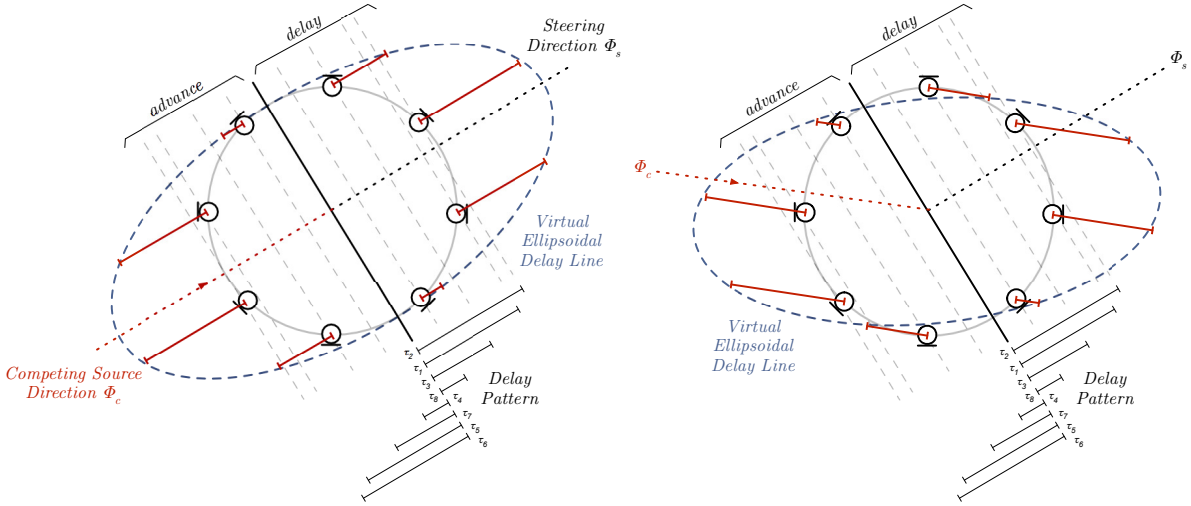


Figure 3.14: This figure shows a delay pattern for signals captured by all microphones of a competing, undesired source.

3.2.3 Spatial Aliasing and Grating Lobes due to Phase Ambiguity

Spatial aliasing arises when the resulting phase shift in a single microphone-pair–combination (see Fig. 3.15) is identical for two or more different directions for a given frequency. Grating lobes occur when all delay-shifts of (3.12) are multiples of 2π

$$\frac{2\pi}{\lambda} r_n(\cos(\phi_s - \phi_n) - \cos(\varphi - \phi_n)) = 2\pi \cdot m$$

or

$$r_n(\cos(\phi_s - \phi_n) - \cos(\varphi - \phi_n)) = m \cdot \lambda, \ \forall n$$

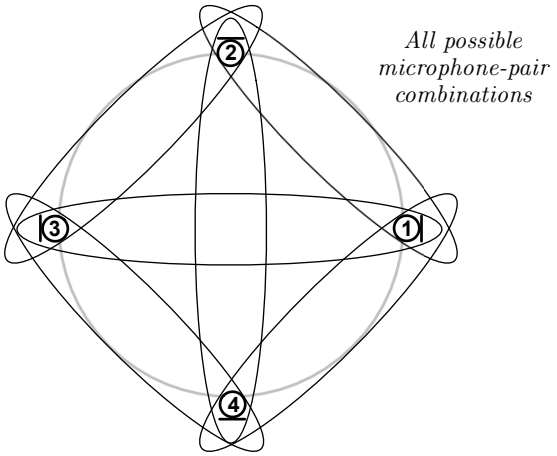


Figure 3.15: This figure shows all possible microphone-pair–combinations of a UCA.

with $m \in \mathbb{N}$.

3.2.4 Characteristics of a Uniform Circular Array

Fig. 3.16 depicts two surface plots, two one-dimensional plots and two polar plots of a beam pattern of a UCA consisting of eight (a, c, e) and twelve (b, d, f) microphones with a diameter of $d = 0.55$m and a steering direction of $\phi_s = 0°$. All plots exhibit the same properties as mentioned in case of a ULA. The DS-BF is considered in all scenarios.

A close look at the first two plots (a) and (b) reveals that the main lobe of the beam pattern is clearly visible and narrow. It exhibits a cardioid characteristic at lower frequencies ($100 - 1000$ Hz). At first sight, a single side lobe (a) and two side lobes (b)—both change the peak position and the width continuously over frequency—are recognizable at lower frequencies. The area outside of the recognizable side lobes consists of an arbitrary lobe pattern which is symmetric around the main response axis. It consists of areas with high (bright) and low (dark) attenuation.

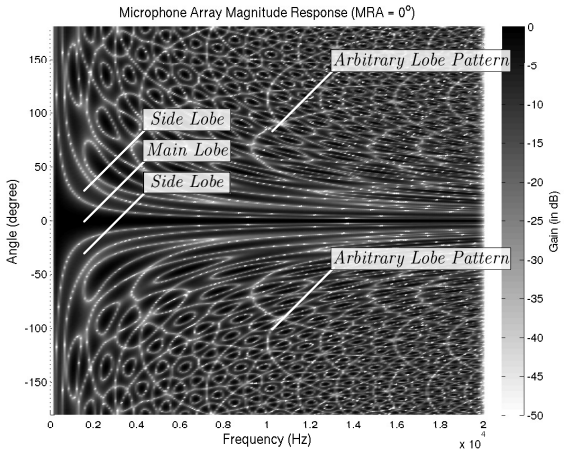
Fig. 3.17 and Fig. 3.18 depicts the same type of plots mentioned above for a UCA consisting of 16 (a, c, e) and 24 (b, d, f) microphones, but with a diameter of $d = 0.55$m and $d = 0.20$m (in this order) and a steering direction of $\phi_s = 0°$. An increase in microphones leads to

- an increased number of side lobes, which change the peak position and the width continuously over frequency, and
- an increased number of arbitrary areas mentioned above.

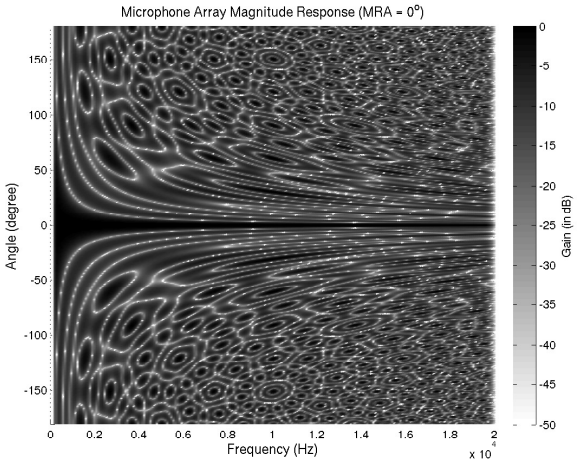
A decrease in diameter yields

- an increased main and side lobe width, and
- a decreased number of arbitrary areas.

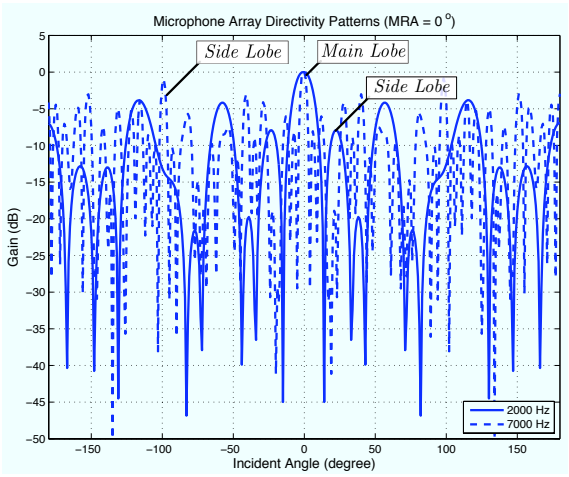
A shift of the steering direction ϕ_s as shown in Fig. 3.19 does not lead to serious changes in the beam pattern at lower frequencies, but significant changes at higher frequencies, and small changes in the width and height of the main and side lobes. It is interesting to see that the main lobe of the UCA is much smaller in contrast to the main lobe of the ULA, which is clearly visible at very low and high frequencies. The higher the diameter, the higher the number of peaks and valleys in the arbitrary lobe pattern. An attenuation of 0 dB is not reached in the audible range, although these areas may exhibit a very low attenuation.



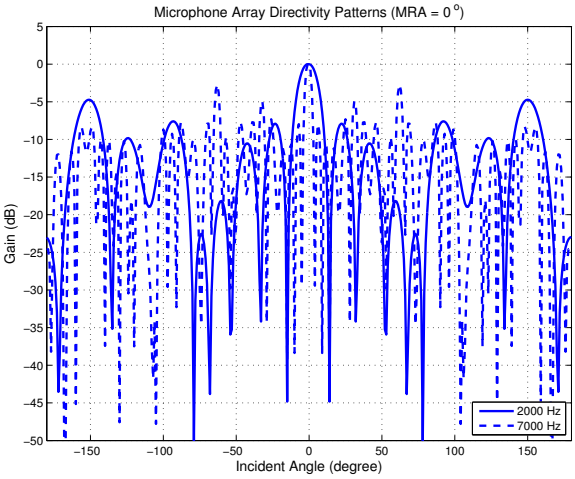
(a) Surface plot of an eight-element array.



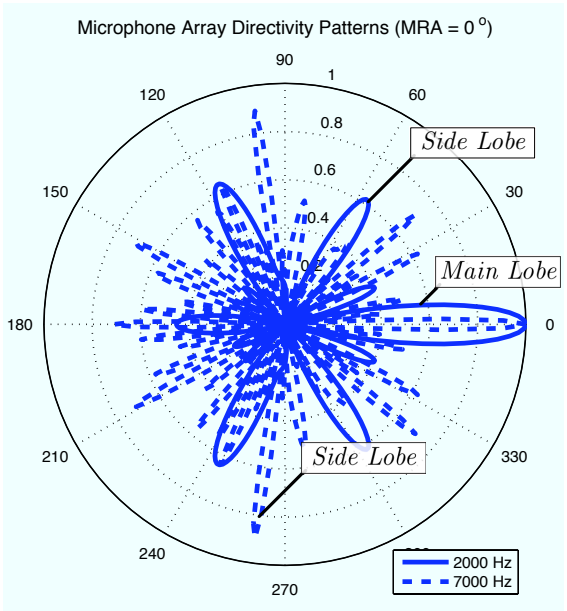
(b) Surface plot of a twelve-element array.



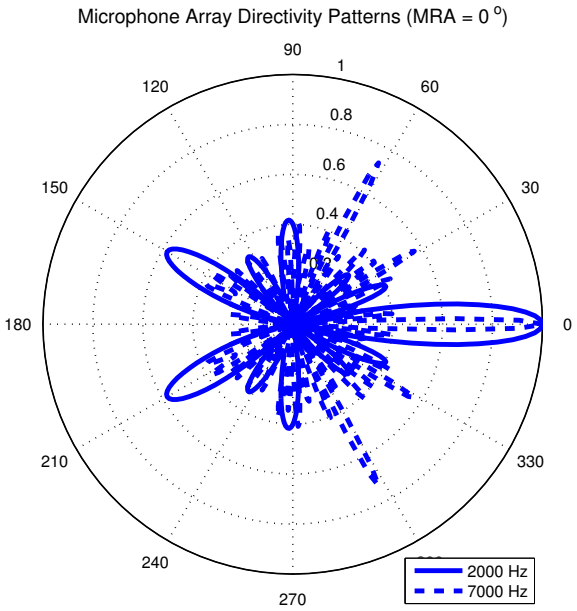
(c) One-dimensional plot of a eight-element array.



(d) One-dimensional plot of a twelve-element array.

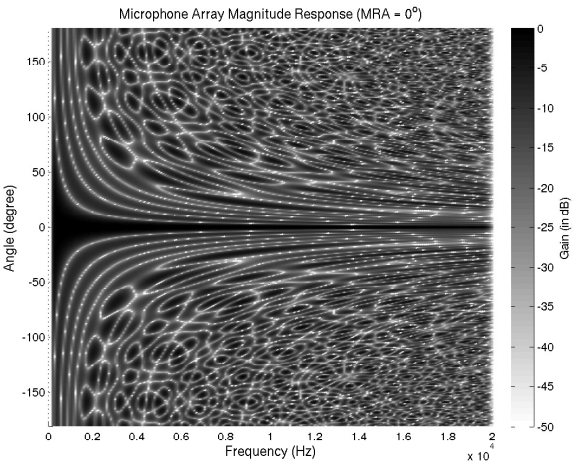


(e) Polar plot of a eight-element array.

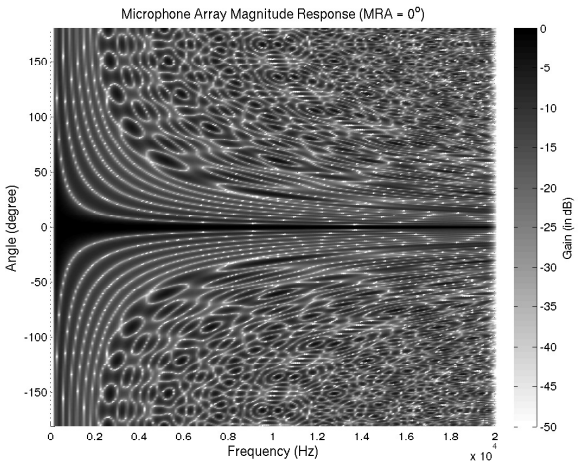


(f) Polar plot of a twelve-element array.

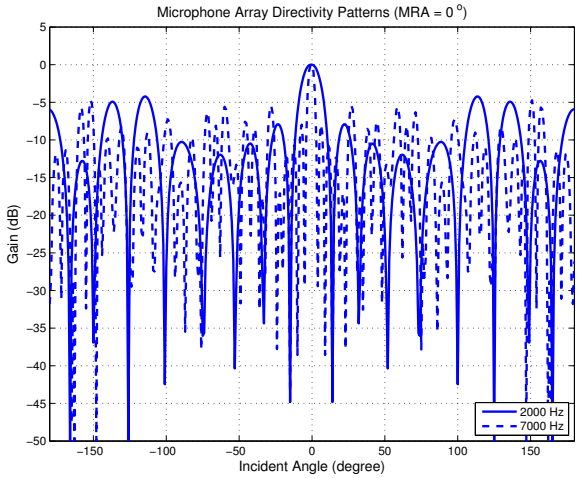
Figure 3.16: The surface plots (a,b) show the beam pattern for all frequencies and angles. The one-dimensional plots show the beam pattern for the given frequencies $f = \{2000, 7000\}$ Hz, and so are the polar plots (e-f). Computations are based on a UCA consisting of eight (a,c,e) and twelve (b,d,f) microphones. The diameter of the UCA is $d = 0.55$ m.



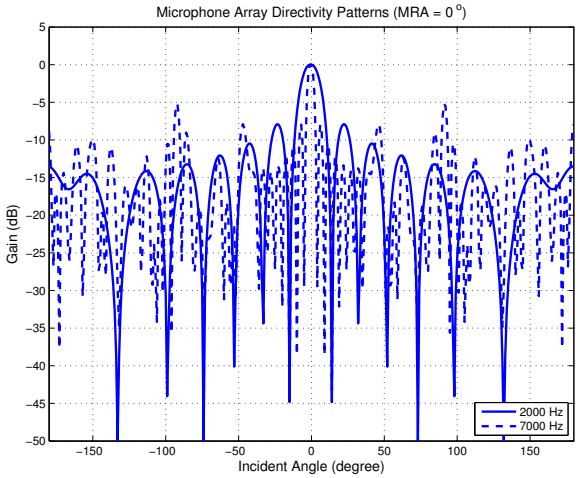
(a) Surface plot of a 16-element array.



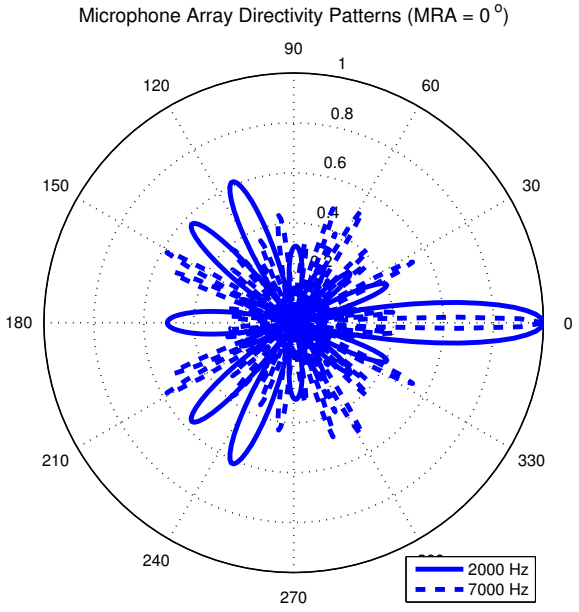
(b) Surface plot of a 24-element array.



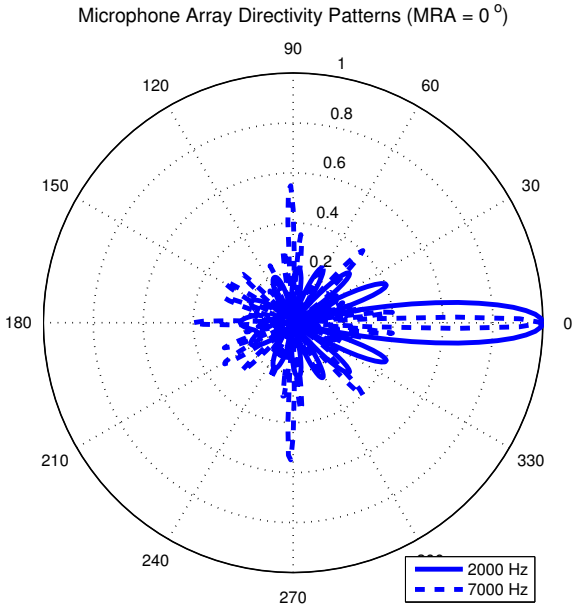
(c) One-dimensional plot of a 16-element array.



(d) One-dimensional plot of a 24-element array.

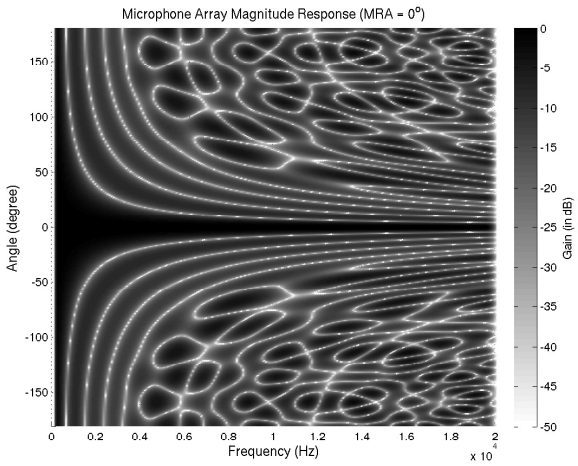


(e) Polar plot of a 16-element array.

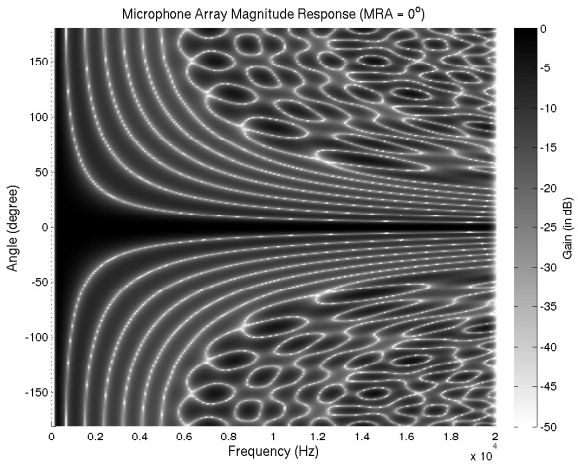


(f) Polar plot of a 24-element array.

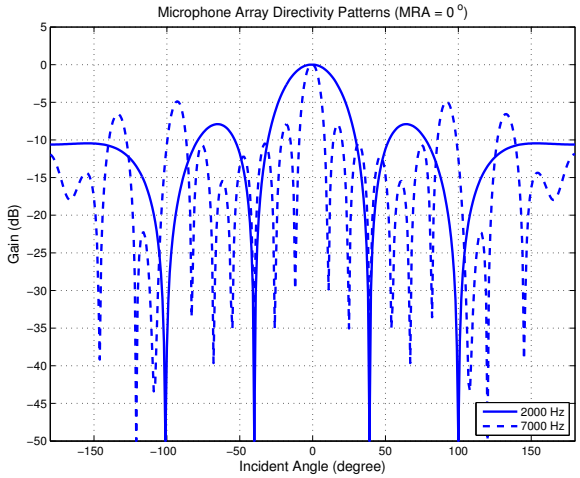
Figure 3.17: The surface plots (a,b) show the beam pattern for all frequencies and angles. The one-dimensional plots show the beam pattern for the given frequencies $f = \{2000, 7000\}$ Hz, and so are the polar plots (e-f). Computations are based on a UCA consisting of 16 (a,c,e) and 24 (b,d,f) microphones. The diameter of the UCA is $d = 0.55$ m.



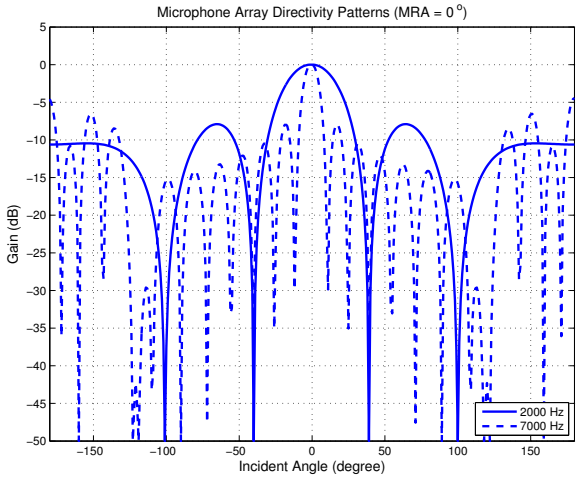
(a) Surface plot of a 16-element array.



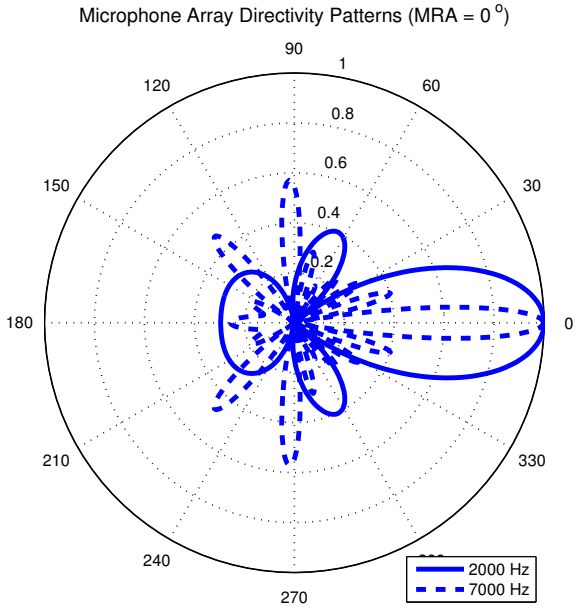
(b) Surface plot of a 24-element array.



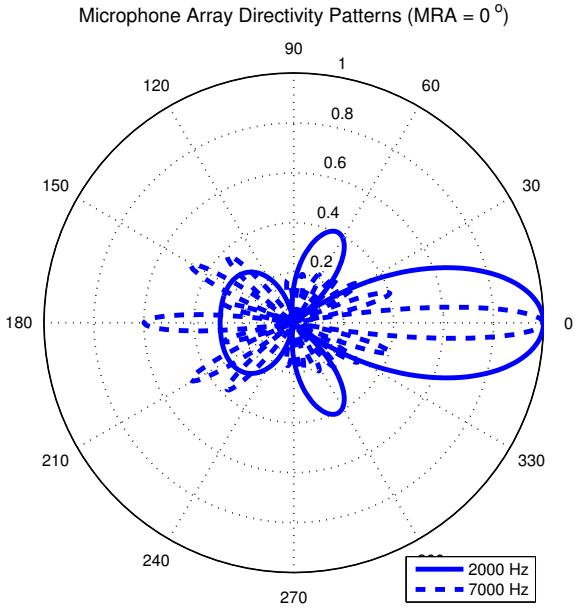
(c) One-dimensional plot of a 16-element array.



(d) One-dimensional plot of a 24-element array.

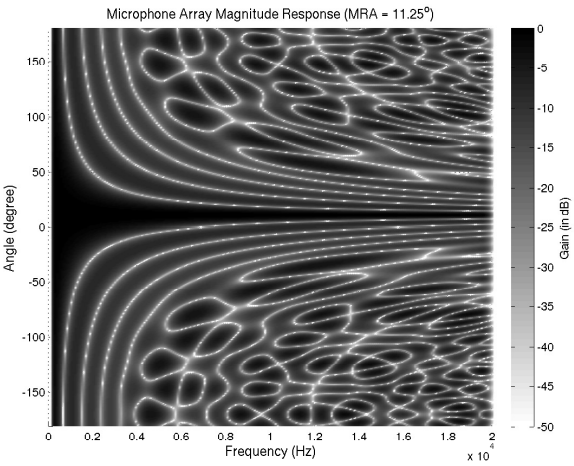


(e) Polar plot of a 16-element array.

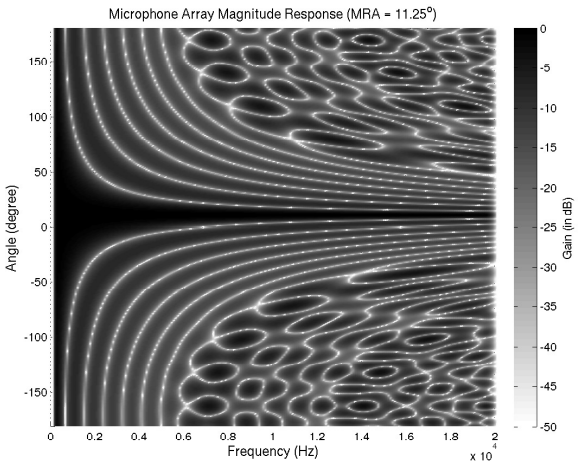


(f) Polar plot of a 24-element array.

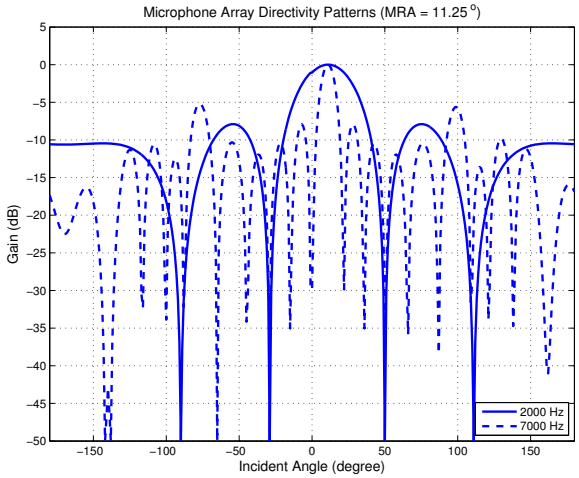
Figure 3.18: The surface plots (a,b) show the beam pattern for all frequencies and angles. The one-dimensional plots show the beam pattern for the given frequencies $f = \{2000, 7000\}$ Hz, and so are the polar plots (e-f). Computations are based on a UCA consisting of 16 (a,c,e) and 24 (b,d,f) microphones. The diameter of the UCA is $d = 0.20$ m.



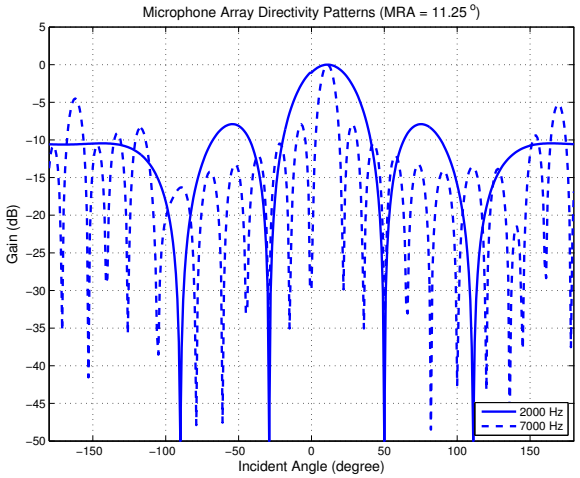
(a) Surface plot of a 16-element array.



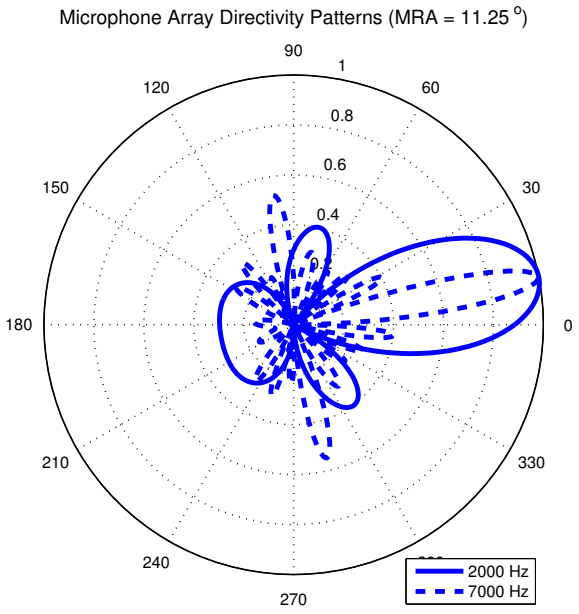
(b) Surface plot of a 24-element array.



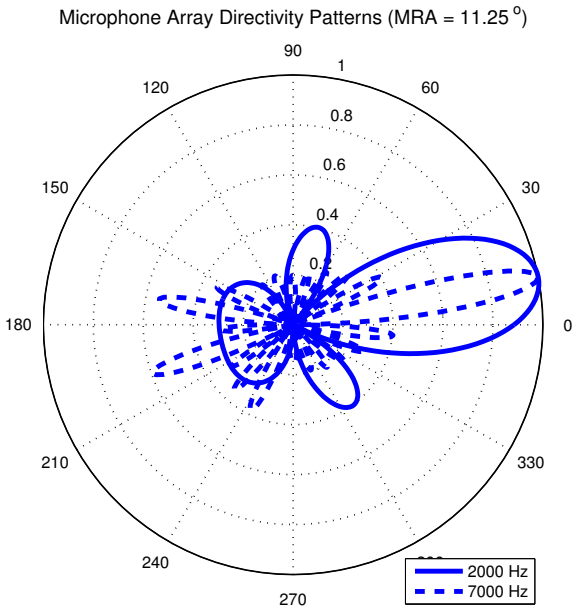
(c) One-dimensional plot of a 16-element array.



(d) One-dimensional plot of a 24-element array.



(e) Polar plot of a 16-element array.



(f) Polar plot of a 24-element array.

Figure 3.19: In this figure, computations are based on a UCA consisting of 16 (a,c,e) and 24 (b,d,f) microphones. The diameter of the UCA is $d = 0.20$ m and—in comparison to the previous figures—the steering direction is set to 11.25°.

4 Beamforming with Uniform Circular Arrays

4.1 Beamforming

Beamforming is a special technique in signal processing which enables source localization, source separation, signal de-reverberation, etc. It refers to designing a spatio-temporal filter [4], i.e. it manipulates signals in temporal and spatial domain.

Captured signals contain signal components from different sources—a desired source and competing sources. Temporal filtering only doesn't work well in case of eliminating the competing sources, because the desired and competing sources may occupy the same frequency bands, and filtering the affected bands leads to a loss of the desired-source–information. A beamformer extracts a desired signal from a specific direction from a reverberant environment influenced by interfering sources. It combines the signals captured by the microphones in a way that signals from a certain direction experience constructive overlapping while others experience destructive interference. The beamformer modifies the directionality of the array.

The use of beamforming algorithms depends on the characteristics of the environment. For instance, a conference room may exhibit a strong/weak presence of reverberation, it may exhibit interfering sources such as a video projector, a (CPU) fan, an air conditioner, etc. On the one hand a DS-BF is able to filter out stationary, non-coherent noise signals, but on the other hand its performance decreases if there is a lot of reverberation, and if the noise signals or interferences are coherent.

4.2 Steering Delay Quantization

The effect of steering delay quantization in case of discrete time systems is rarely discussed in literature, although it may influence the beam pattern negatively. It leads to a loss of the beam pattern symmetry, which strongly depends on the sampling frequency of the ADC, and it affects the results if the ADC works at the Nyquist rate of the desired signal[19] [11].

[19] The Nyquist rate is exactly twice the bandwidth of a bandlimited signal.

Let's consider a delay mentioned in the previous sections:

$$\boxed{\tau_n = \frac{r_n \cdot \cos(\varphi - \phi_n)}{c}} \,. \tag{4.1}$$

It describes the delay generated by the beamformer for the waves captured by the microphone with index n. These delays steer the beamformer into the direction of the desired source in case of $\varphi = \phi_s$. In time-domain signal processing, they have to be integer multiples of the sampling period. If the delay is fractional, it can be rounded to a delay closest to the fractional delay; but this causes little changes in the beam pattern. Another way to allow fractional delays is to increase the sampling rate of the ADC, or by considering up-sampling and interpolation, time-shifting and down-sampling during signal processing. Nevertheless, this method needs much more resources than rounding.

A better way to consider fractional delays is doing these temporal shifts in frequency domain. In case of block-processing, a DFT or FFT transforms the time-domain signal into frequency domain. An arbitrary phase shift in frequency domain leads to a change of the phase spectrum. A transformation back into time domain yields a signal with modified amplitudes that correspond to the amplitudes of the fractional shifted signal. In this case, the steering delay quantization error depends only on the amplitude resolution of the system, which is generally high enough for a 16-bit quantizer and higher. Fig. 4.1 shows the steering delay quantization with a low amplitude resolution. A small phase shift in frequency domain does not lead to a time shift in time domain because of a bad amplitude resolution.

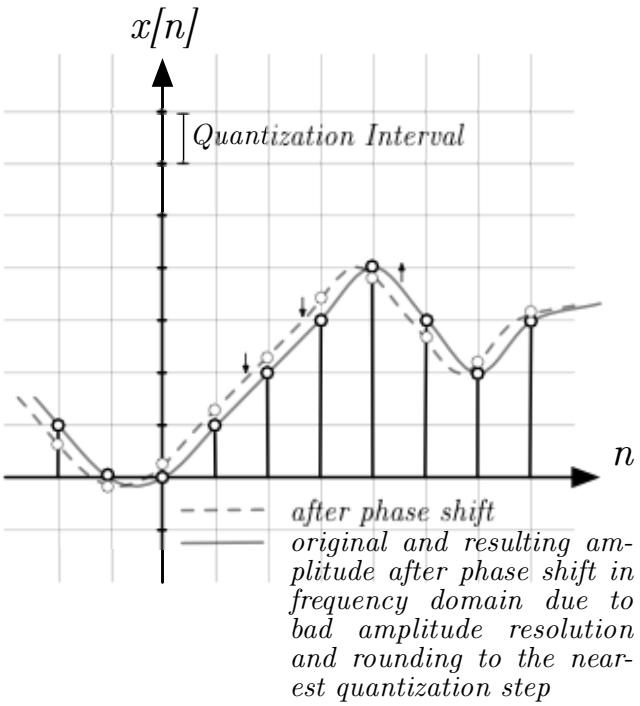


Figure 4.1: This figure shows the steering delay quantization with a low amplitude resolution and rounding to the nearest quantization step. A phase shift in frequency domain does not lead to any shift in time domain.

4.3 Delay&Sum Beamformer

4.3.1 The Algorithm

The DS-BF is a data-independent beamformer [12]. Its coefficients for different steering directions ϕ_s can be calculated offline. Thus, a DSP can access a database consisting of all coefficients instead of calculating the coefficients in real time.

The main tasks of a DS-BF are compensating the relative delays between the captured signals [7], summarizing the shifted signals and scaling the sum with the number of microphones to avoid a signal N times bigger than the amplitude of each captured signal in a far-field condition. The shifts depend on the radius of the UCA, the position of all microphones $\boldsymbol{r}_n$, the sound velocity c, and the steering direction ϕ_s.

The algorithm is efficient in case of non-coherent noise sources, i.e. spatially white noise, for a high number of microphones. It is inefficient in case of reverberant environments and coherent noise sources, whereas its performance depends on the time of arrival of the noise source [13].

Fig. 4.2 depicts two capturing scenarios. The left one does not include any beamformer but the right one. In the first scenario the captured signals—each signal contains a single pulse—do not overlap constructively at the output because the pulse arrives at different instants of time at the microphones. Consequently, the captured energy is spread over a long time interval. This may cause a comb-filter–effect and an acoustic smearing of the signal. In the second scenario the captured signals are delayed first, followed by a summation of all signals and a scaling of the resulting sum. The beamformer emphasizes the signals from the steering direction ϕ_s and attenuates signals from other directions.

4.3.2 The Implementation

In this work the coefficients are calculated in frequency domain (see Fig. 4.3) according to

$$\boxed{W_n(\omega) = e^{-i\frac{\omega}{c}\frac{d}{2}\cos(\varphi-\phi_n)}} \tag{4.2}$$

for a single microphone or

$$\boxed{W(\omega) = \frac{1}{N}\sum_{n=1}^{N} e^{-i\frac{\omega}{c}\frac{d}{2}\cos(\varphi-\phi_n)}} \tag{4.3}$$

for the whole array, where ω is the angular frequency, N is the number of microphones, c is the speed of sound, d is the array diameter, ϕ_n is the angle of the microphone with index n, and φ is an arbitrary angle which should be the direction of the desired source. In the following pseudo code N_φ is the number of beams, N_b is the number of frequencies, N_m is the number of microphones, and $\boldsymbol{\phi_n}$ is the microphone angle vector.

Algorithm 1 Delay&Sum Beamformer

```
for j = 1 : N_φ do
    for k = 1 : N_f do
        W(j, :, k) = (1/N_m) e^{-i (2πf(k)/c) (d/2) cos(φ(j) − φ_n)}    ▷ W is a (N_φ × N_m × N_f)-matrix.
    end for
end for
```

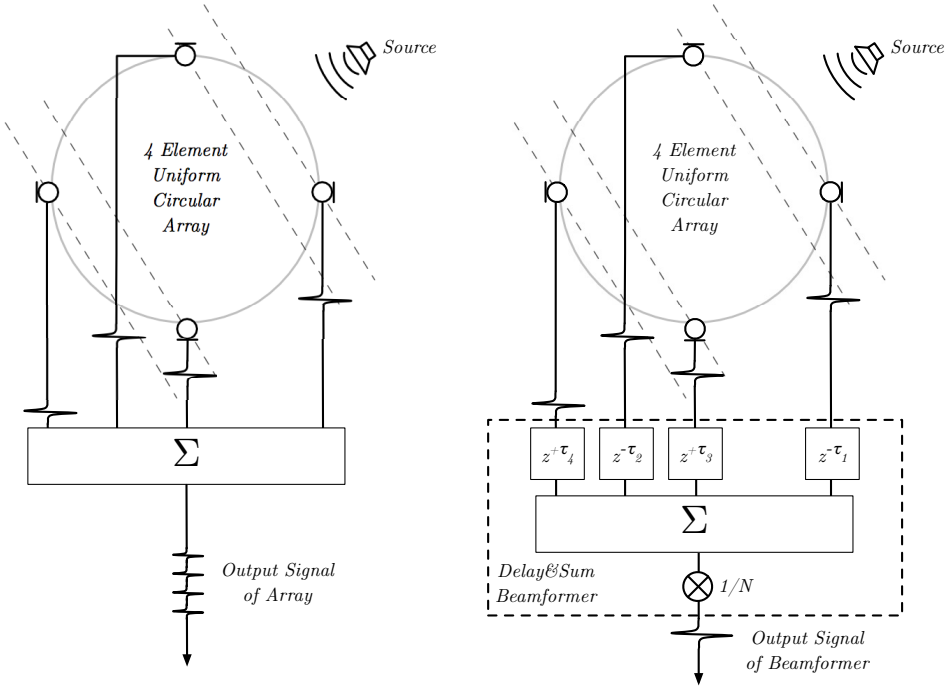


Figure 4.2: Left: Array processing without a beamformer. The output signal consists of the non-delayed captured signals; there is no constructive interference. Right: Array Processing with a DS-BF. The output signal consists of a single pulse due to the compensation of the relative delays.

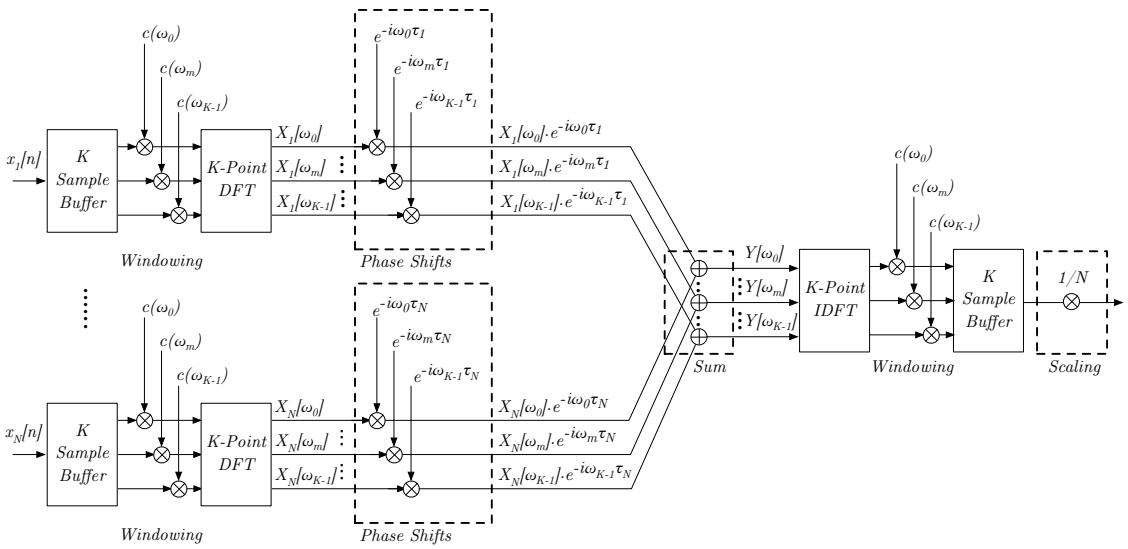


Figure 4.3: Block diagram of the DS-BF.

4.4 Minimum Power Distortionless Response Beamformer with Loading Level and Sample Matrix Inversion

4.4.1 The Algorithm

The MPDR-BF is a data-dependent beamformer [7] which is based on a constrained optimal beamformer design. It is able to emphasize the signals from steering direction ϕ_s and attenuates interfering signals from different directions. The beamformer is sensitive to wrong positioned microphones and calibration errors which results in amplitude and phase deviations in each channel. The optimal weights of the MPDR-BF follows the derivation of the MVDR-BF[20]. The model of the captured signal in frequency domain is

$$\boldsymbol{x}(\omega) = \boldsymbol{d}(\omega)s(\omega) + \boldsymbol{n}(\omega) \tag{4.4}$$

where $\boldsymbol{x}(\omega) = (x_1(\omega), x_2(\omega), .., x_N(\omega))^T$ is the input vector, $\boldsymbol{d}(\omega) = (d_1(\omega), d_2(\omega), .., d_N(\omega))^T$ is the capturing or steering vector, $s(\omega)$ is the desired signal and $\boldsymbol{n}(\omega) = (n_1(\omega), n_2(\omega), .., n_N(\omega))^T$ is the noise vector, and all vectors exhibit the dimension $(N \times 1)$ where N is the number of microphones. The output signal is

$$y(\omega) = \boldsymbol{w}^H(\omega)\boldsymbol{x}(\omega) \tag{4.5}$$

where $(\cdot)^H$ stands for the Hermitian transpose. The main target is to output the desired signal only without any influence of noise and other interferences; that is

$$y(\omega) \stackrel{!}{=} s(\omega) = \boldsymbol{w}^H(\omega)\boldsymbol{x}(\omega) = \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega)s(\omega)$$

which requires $\boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = 1$. Considering noise yields

$$y(\omega) = \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega)s(\omega) + \boldsymbol{w}^H(\omega)\boldsymbol{n}(\omega) = s(\omega) + y_n(\omega).$$

The next step is to minimize the noise variance

$$\mathrm{E}\{|y_n(\omega)|^2\} = \boldsymbol{w}^H(\omega)\mathrm{E}\{\boldsymbol{n}(\omega)\boldsymbol{n}(\omega)^H\}\,\boldsymbol{w}(\omega) \tag{4.6}$$

where $\mathrm{E}\{\cdot\}$ is the expectation operator and $\mathrm{E}\{\boldsymbol{n}(\omega)\boldsymbol{n}(\omega)^H\} = \boldsymbol{R}_{nn}(\omega)$ is the noise cross-power spectral matrix, also known as the array covariance matrix. In case of the MPDR-BF the cross-power matrix of the input signal R_{xx} replaces R_{nn}.

Simple MPDR Beamformer

For an ordinary MPDR-BF the constrained minimization problem according to [14] is

$$\arg\min_{\boldsymbol{w}(\omega)} \boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}(\omega)\boldsymbol{w}(\omega) \tag{4.7}$$

$$\text{subject to } \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = 1\ . \tag{4.8}$$

This optimization problem can be solved by using Lagrange multipliers.

$$J(\boldsymbol{w}, \lambda) = \boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}(\omega)\boldsymbol{w}(\omega) + \lambda(\boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) - 1) + (\lambda[\boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) - 1])^*$$

[20] MVDR - Minimum Variance Distortionless Response Beamformer

which is

$$J(\boldsymbol{w},\lambda) = \boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}(\omega)\boldsymbol{w}(\omega) + \lambda(\boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) - 1) + (\lambda^*[\boldsymbol{d}^H(\omega)\boldsymbol{w}(\omega) - 1]) \tag{4.9}$$

or

$$J(\boldsymbol{w},\lambda) = \boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}(\omega)\boldsymbol{w}(\omega) + 2\mathrm{Re}\big\{\lambda\left(\boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) - 1\right)\big\} \tag{4.10}$$

The complex gradient of (4.9) with respect to $\boldsymbol{w^H}(\omega)$ and assuming $\boldsymbol{w}(\omega)$ as a constant leads to

$$\begin{aligned}&\nabla_{\boldsymbol{w^H}(\omega)} J(\boldsymbol{w},\lambda) = \boldsymbol{R}_{xx}(\omega)\boldsymbol{w}(\omega) + \lambda\boldsymbol{d}(\omega) = 0 \;/\cdot \boldsymbol{R}_{xx}^{-1}(\omega)\\ &0 = \boldsymbol{w}(\omega) + \boldsymbol{R}_{xx}^{-1}(\omega)\lambda\boldsymbol{d}(\omega)\end{aligned}$$

and

$$\boldsymbol{w}(\omega) = \boldsymbol{R}_{xx}^{-1}(\omega)\,(-\lambda)\,\boldsymbol{d}(\omega) \tag{4.11}$$

Considering $\boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = \boldsymbol{d}^H(\omega)\boldsymbol{w}(\omega) = 1$ yields

$$\boldsymbol{d}^H(\omega)\boldsymbol{w}(\omega) = 1 = \boldsymbol{d}^H(\omega)\boldsymbol{R}_{xx}^{-1}(\omega)\,(-\lambda)\,\boldsymbol{d}(\omega). \tag{4.12}$$

Thus, $(-\lambda)$ has to be $\boldsymbol{d}^H(\omega)\boldsymbol{R}_{xx}^{-1}(\omega)\boldsymbol{d}(\omega)$ to satisfy (4.12). Finally, this results in

$$\boxed{\boldsymbol{w}(\omega) = \frac{\boldsymbol{R}_{xx}^{-1}(\omega)\boldsymbol{d}(\omega)}{\boldsymbol{d}^H(\omega)\boldsymbol{R}_{xx}^{-1}(\omega)\boldsymbol{d}(\omega)}}\,. \tag{4.13}$$

Advanced MPDR Beamformer with Diagonal Loading

As mentioned before, the standard MPDR-BF exhibits bad performance in case of different mismatches. An additional quadratic constraint increases its performance [15]. The optimization is modified as

$$\arg\min_{\boldsymbol{w}(\omega)} \boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}(\omega)\boldsymbol{w}(\omega) \tag{4.14}$$

$$\text{subject to } \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = 1,\ \|\boldsymbol{w}(\omega)\|^2 \leq T\ , \tag{4.15}$$

which leads to

$$\boxed{\boldsymbol{w}_{DL}(\omega) = \frac{\left(\boldsymbol{R}_{xx}(\omega) + \gamma\boldsymbol{I}\right)^{-1}\boldsymbol{d}(\omega)}{\boldsymbol{d}^H(\omega)\left(\boldsymbol{R}_{xx}(\omega) + \gamma\boldsymbol{I}\right)^{-1}\boldsymbol{d}(\omega)}} \tag{4.16}$$

where T is the quadratic norm threshold, γ is the loading level, and $\boldsymbol{I}$ is the identity matrix. According to [15] and [16], the increase in robustness of the MVDR-BF and MPDR-BF is a trade-off between the suppression of the side lobes and the ability to cancel interferences and attenuate noise. If the loading level γ is zero, the beamformer behaves as an ordinary but sensitive MPDR-BF. If $\gamma = \infty$, the beamformer behaves as a DS-BF[17]. The use of Newton's method may lead to the optimal loading level parameters, but this requires the knowledge of all imbalances.

Advanced MPDR Beamformer with Variable Loading

Another way and a more general approach is modifying the additional constraint according to [15]:

$$\arg\min_{\boldsymbol{w}(\omega)} \boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}(\omega)\boldsymbol{w}(\omega) \tag{4.17}$$

$$\text{subject to } \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = 1,\ \boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}^{-1}(\omega)\boldsymbol{w}^H(\omega) \leq T\ , \tag{4.18}$$

which leads to variable loading levels for the eigenvalues of $\boldsymbol{R}_{xx}(\omega)$ and the weighting coefficients

$$\boxed{\boldsymbol{w}_{VL}(\omega) = \frac{\left(\boldsymbol{R}_{xx}(\omega) + \delta\boldsymbol{R}_{xx}^{-1}(\omega)\right)^{-1}\boldsymbol{d}(\omega)}{\boldsymbol{d}^H(\omega)\left(\boldsymbol{R}_{xx}(\omega) + \delta\boldsymbol{R}_{xx}^{-1}(\omega)\right)^{-1}\boldsymbol{d}(\omega)}} \tag{4.19}$$

where δ satisfies the condition $\boldsymbol{w}^H(\omega)\boldsymbol{R}_{xx}^{-1}(\omega)\boldsymbol{w}^H(\omega) \leq T$. This implementation yields a better robustness without loosing much adaptivity in comparison to the previous MPDR-BF implementations.

4.4.2 The Implementation

The biggest problem in the real-time implementation is the estimation of the correlation matrix which results in a sample covariance matrix $\hat{\boldsymbol{R}}_{xx}(\omega)$ and a sample covariance matrix inversion $\hat{\boldsymbol{R}}_{xx}^{-1}(\omega)$ [18]. One possible implementation of a covariance matrix estimator is

$$\hat{\boldsymbol{R}}_{xx}(\omega) = \frac{1}{K}\sum_{n=0}^{K-1}\boldsymbol{x}[n]\boldsymbol{x}[n]^H \tag{4.20}$$

where $\boldsymbol{x}[n]\boldsymbol{x}[n]^H$ yields an $(N \times N)$-matrix for each time step n, and K scales the sum of these matrices. The estimator considers a rectangular window function. The estimated matrix can be ill-conditioned or inaccurately estimated because of a lack of training data, silent signals, or non-stationary interferences. Diagonal or variable loading with proper loading level values eliminate the problem of bad-conditioned matrices.

In this work the coefficients are calculated in frequency domain in two different ways. The first implementation considers diagonal loading according to

$$\boxed{\boldsymbol{w}_{DL}(\omega) = \frac{\left(\boldsymbol{R}_{xx}(\omega) + \gamma\boldsymbol{I}\right)^{-1}\boldsymbol{d}(\omega)}{\boldsymbol{d}^H(\omega)\left(\boldsymbol{R}_{xx}(\omega) + \gamma\boldsymbol{I}\right)^{-1}\boldsymbol{d}(\omega)}} \tag{4.21}$$

where $\gamma = \boldsymbol{x}(\omega)^H\boldsymbol{x}(\omega)\cdot 10^{-3}$, and the second implementation computes its coefficients according to

$$\boxed{\boldsymbol{w}_{VL}(\omega) = \frac{\left(\boldsymbol{R}_{xx}(\omega) + \delta\boldsymbol{R}_{xx}^{-1}(\omega)\right)^{-1}\boldsymbol{d}(\omega)}{\boldsymbol{d}^H(\omega)\left(\boldsymbol{R}_{xx}(\omega) + \delta\boldsymbol{R}_{xx}^{-1}(\omega)\right)^{-1}\boldsymbol{d}(\omega)}} \tag{4.22}$$

where $\delta = 10^{-2}$ [15] (Note: The values of γ and δ depend on the signals (speech or music) and other conditions (e.g., single- or double-talk)). The spectrum of each captured signal frame is weighted with the corresponding coefficients (see Fig. 4.4) after calculating the coefficients. The pseudo code of the diagonal and variable loading algorithms is shown in Algorithm 2 and 3. In the following pseudo code N_φ is the number of beams, N_b is the number of frequencies, N_m is the number of microphones, and $\boldsymbol{\phi_n}$ is the microphone position vector.

Algorithm 2 Minimum Power Distortionless Response Beamformer with Diagonal Loading

$\gamma = \boldsymbol{x}^H(\omega)\boldsymbol{x}(\omega) \cdot a_4$ ▷ γ is the loading level.
$\lambda = a_5$ ▷ λ is the forgetting factor.
for $j = 1 : N_\varphi$ **do**
  for $k = 1 : N_f$ **do**
    $\boldsymbol{d}(:) = e^{i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi_n})}$ ▷ $\boldsymbol{d}$ is a $(N_m \times 1)$-steering-vector.
    $\boldsymbol{R}_{xx} = \lambda \cdot \boldsymbol{R}_{xx} + (1-\lambda) \cdot \boldsymbol{n}(\omega)\boldsymbol{n}^H(\omega) + \gamma \boldsymbol{I}$ ▷ $\boldsymbol{R}_{xx}$ is the covariance matrix.
    $\boldsymbol{W}(j,:,k) = \frac{\boldsymbol{R}_{xx}^{-1}\cdot\boldsymbol{d}}{\boldsymbol{d}^H\boldsymbol{R}_{xx}^{-1}\cdot\boldsymbol{d}}$ ▷ $\boldsymbol{W}$ is a $(N_\varphi \times N_m \times N_f)$-matrix.
  end for
end for

Algorithm 3 Minimum Power Distortionless Response Beamformer with Variable Loading

$\delta = a_4$ ▷ δ is the loading level.
$\lambda = a_5$ ▷ λ is the forgetting factor.
$\epsilon = a_6$ ▷ Use ϵ to avoid badly scaled matrix.
for $j = 1 : N_\varphi$ **do**
  for $k = 1 : N_f$ **do**
    $\boldsymbol{d}(:) = e^{i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi_n})}$ ▷ $\boldsymbol{d}$ is a $(N_m \times 1)$-steering-vector.
    $\boldsymbol{R}_{xx} = \lambda \cdot \boldsymbol{R}_{xx} + (1-\lambda) \cdot \boldsymbol{n}(\omega)\boldsymbol{n}^H(\omega) + \epsilon \boldsymbol{I}$ ▷ $\boldsymbol{R}_{xx}$ is the covariance matrix.
    $\boldsymbol{W}(j,:,k) = \frac{\left(\boldsymbol{R}_{xx}+\delta\boldsymbol{R}_{xx}^{-1}\right)^{-1}\boldsymbol{d}}{\boldsymbol{d}^H\left(\boldsymbol{R}_{xx}+\delta\boldsymbol{R}_{xx}^{-1}\right)^{-1}\boldsymbol{d}}$ ▷ $\boldsymbol{W}$ is a $(N_\varphi \times N_m \times N_f)$-matrix.
  end for
end for

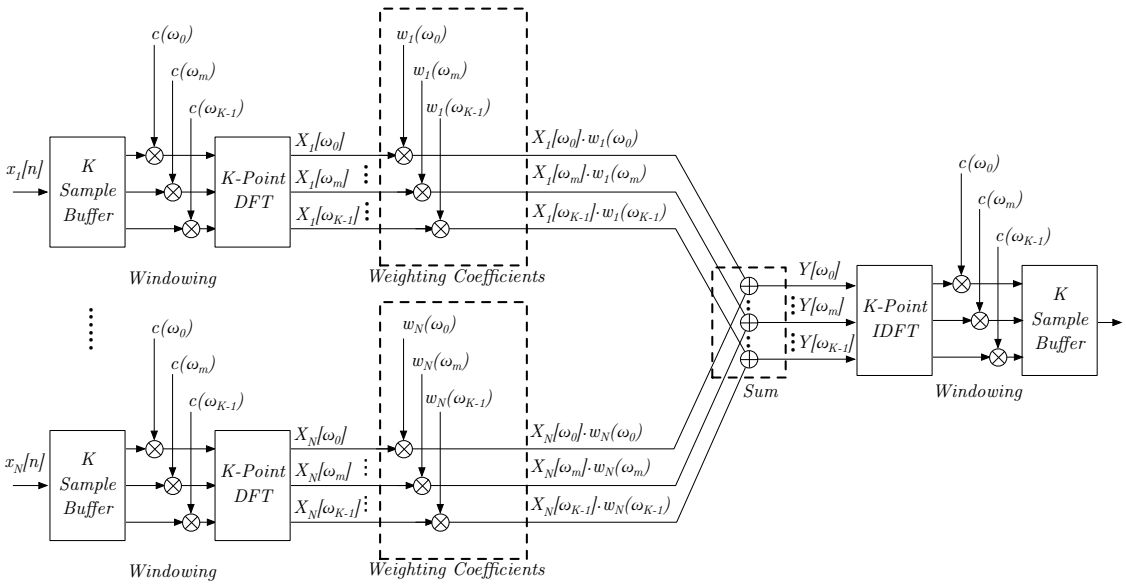


Figure 4.4: Block diagram of the MPDR-BF.

4.5 Robust Least Squares Frequency Invariant Beamformer

4.5.1 The Algorithm

The RLSFI-BF[21] is a data-independent beamformer [19]. The aim of this super-directive beamformer is to focus the desired source with a beam exhibiting a small and constant beamwidth—especially at lower frequencies—and very small side lobes. The DS-BF features a flat beam pattern at lower frequencies over all angles, whereas the RLSFI-BF exhibits a far better behaviour at lower frequencies, and it entails a higher directivity index [20] using a small array aperture [19]. It is 'extremely' sensitive to spatially white noise and sensor mismatches, e.g., position errors and deviations in the microphone characteristics. The algorithm, which computes the beamformer coefficients, considers the white noise gain—it gives information about the ability to suppress spatially white noise—and the undistorted signal response from the desired looking direction as optimization constraints. The algorithm is based on convex optimization methods.

4.5.2 The Implementation

The beamformer design is based on least squares computations that approximate a given response numerically and in the sense of least squares. The performance of the resulting beamformer strongly depends on the angle and frequency resolution. The desired response is defined as follows:

$$\hat{b}(f,\varphi) = \sum_{n=1}^{N} w_n(f) e^{-i\frac{2\pi f}{c}\frac{d}{2}\cos(\varphi-\phi_n)} \tag{4.23}$$

or in vector notation

$$\hat{\boldsymbol{B}}(f) = \boldsymbol{G}(f)\boldsymbol{w}(f) \tag{4.24}$$

where $\boldsymbol{w}(f)^T = (w_1(f), w_2(f), ..., w_N(f))$ is the coefficient vector and $\boldsymbol{G}(f)$ is a matrix containing $M \times N$ elements according to $G_{m,n} = e^{-i\frac{2\pi f}{c}\frac{d}{2}\cos(\varphi_m-\phi_n)}$. A simple LS-solution is obtained by minimizing

$$\arg\min_{\boldsymbol{w}(f)} \|\boldsymbol{G}(f)\boldsymbol{w}(f) - \hat{\boldsymbol{B}}(f)\|_2^2 \tag{4.25}$$

$$\text{subject to } \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = 1 \tag{4.26}$$

where $\boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = 1$ is the constraint for an undistorted desired signal, and $\|\cdot\|_2$ is the L_2-norm. In general, $\|\cdot\|_p$ is the L_p-norm which is defined as

$$\|x\|_p := \sqrt[p]{\sum_{i=1}^{N} |x_i|^p}. \tag{4.27}$$

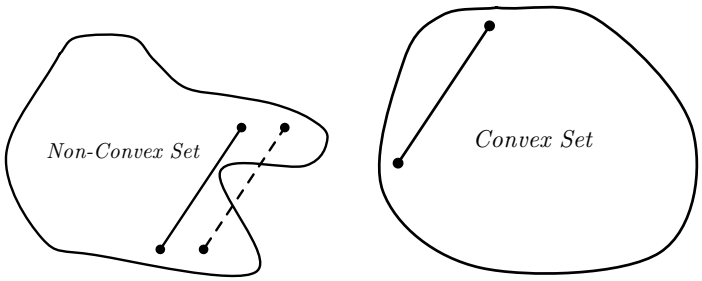
The RLSFI-BF assumes the same desired response for all frequencies, i.e. $\hat{\boldsymbol{B}}(f) = \hat{\boldsymbol{B}}$ and

$$\arg\min_{\boldsymbol{w}(f)} \|\boldsymbol{G}(f)\boldsymbol{w}(f) - \hat{\boldsymbol{B}}(f)\|_2^2 \tag{4.28}$$

$$\text{subject to } \frac{|\boldsymbol{w}^T(f)\boldsymbol{d}(f)|^2}{\boldsymbol{w}^H(f)\boldsymbol{w}(f)} \geq \gamma,\ \boldsymbol{w}^H(\omega)\boldsymbol{d}(\omega) = 1, \tag{4.29}$$

[21] RLSFI-BF - Robust Least Squares Frequency Invariant Beamformer

where the first constraint describes the white noise gain bounded by a lower bound γ. The lower bound is a parameter which enables controlling the robustness of the beamformer [19]. If the gain is smaller than one, it amplifies the spatially white noise. In case of a super-directive beamformer the white noise gain is smaller than 10^{-3} at lower frequencies, and that's the reason why the RLSFI-BF is sensitive to white noise. The unconstrained least-squares problem (4.28) and both constraints (4.29) have to span a convex set in the Euclidean space. All points within this set can be joined with a straight line without leaving this set; a cube or a circle exhibit this property (see Fig. 4.5a). If a line-segment is outside of this set, it's defined as a non-convex set (see Fig. 4.5b), and convex optimization methods are not able to determine the optimal coefficients. According to [21] (4.28) is a convex function because of its quadratic L_2-norm. The constraints are convex too; equation (4.29a) describes an Euclidean ball, whereas the elements of (4.29b) lie in a hyper-plane. If the unconstrained least-squares problem (4.28) and both constraints (4.29) exhibit convexity, convex optimization algorithms are able to approximate the optimal solutions, e.g., by using the modeling system for disciplined convex programming `cvx`[22] which is efficient in case of constrained norm minimization [21]. The pseudo code of the beam-design is shown in Algorithm 4. In the following pseudo code N_φ is the number of beams, N_b is the number of frequencies, N_m is the number of microphones, and $\boldsymbol{\phi_n}$ is the microphone position vector.



(a) A non-convex set. (b) A convex set.

Figure 4.5: In case of a non-convex set there are points which are not connectable with each other point without leaving the set, whereas in case of a convex set every point is connectable with each other point within the set.

Algorithm 4 Robust Least Squares Frequency Invariant Beamformer

$[\sim, l] = min(abs(\boldsymbol{\varphi} - \phi_s))$
$\hat{\boldsymbol{B}}(l) = 1/\sqrt{2}$
for $k = 1 : N_f$ **do**
  for $j = 1 : N_\varphi$ **do**
    $\boldsymbol{G}(j, :, k) = e^{-i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi})}$
  end for
  $\boldsymbol{d}(:, k) = e^{i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi_n})}$ ▷ Far-Field model.

  cvx_begin quiet ▷ Start cvx programming.
    variable $w(N_m)$ complex
      minimize $\|\boldsymbol{G}(:, :, k)\boldsymbol{w}(k) - \hat{\boldsymbol{B}}(:)\|_2^2$
    subject to
      $\boldsymbol{w}^H\boldsymbol{d} == 1$
      $\boldsymbol{w}^H\boldsymbol{w} <= 1/\gamma$
  cvx_end ▷ End cvx programming.
end for

[22] cvx: `http://cvxr.com/cvx/`

4.6 Multiple Null Synthesis Robust Least Squares Frequency Invariant Beamformer

4.6.1 The Algorithm

The MNS-RLSFI-BF[23] is a data-independent beamformer based on convex optimization methods [19] and the RLSFI-BF, but with different parameters and an additional constrained mentioned in [22]. It enables multiple null-placement in different directions. Its main lobe is slightly broader than the main lobe of the RLSFI-BF for frequencies between 100 Hz to 16000 Hz—it depends on the array geometry—but still smaller than the main lobe of the DS-BF. The beamformer is as sensitive to spatially white noise and sensor mismatches as the RLSFI-BF. Combining both, the RLSFI-BF and the MNS-RLSFI-BF, yields an improvement in performance, if the RLSFI-BF is used at lower, and its modification at higher frequencies.

4.6.2 The Implementation

The beamformer design is based on the design mentioned in Section 4.5, but with an additional constraint according to

$$\arg\min_{\boldsymbol{w}(f)} \|\boldsymbol{G}(f)\boldsymbol{w}(f) - \hat{\boldsymbol{B}}(f)\|_2^2 \tag{4.30}$$

$$\text{subject to } \frac{|\boldsymbol{w}^T(f)\boldsymbol{d}(f)|^2}{\boldsymbol{w}^H(f)\boldsymbol{w}(f)} \geq \gamma,\ \boldsymbol{w}^T(\omega)\boldsymbol{d}(\omega) = 1 \tag{4.31}$$

and

$$\text{subject to } \boldsymbol{w}^H(\omega)\boldsymbol{V}(\omega) = \boldsymbol{0} \tag{4.32}$$

where

$$\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_S] \tag{4.33}$$

is a matrix which consists of vectors $\boldsymbol{v}$ that describe the sound capture model of the competing sources, and S is the number of nulls. Again, convex optimization algorithms determine the weighting coefficients, as shown in Section 4.5. The pseudo code of the new beam-design is shown in Algorithm 5. In the following pseudo code N_φ is the number of beams, N_b is the number of frequencies, N_m is the number of microphones, and $\boldsymbol{\phi_n}$ is the microphone position vector.

[23] MNS-RLSFI-BF - Multiple Null Synthesis Robust Least Squares Frequency Invariant Beamformer

Algorithm 5 Multiple Null Synthesis Robust Least Squares Frequency Invariant Beamformer

$[\sim, l] = min(abs(\boldsymbol{\varphi} - \phi_s))$
$\hat{\boldsymbol{B}}(l) = 1/\sqrt{2}$

for $k = 1 : N_f$ **do**
  for $j = 1 : N_\varphi$ **do**
    $\boldsymbol{G}(j, :, k) = e^{-i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi})}$
  end for

  $\boldsymbol{d}(:, k) = e^{i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi_n})}$
  $\boldsymbol{v}_1(:, k) = e^{i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi_n})}$
  ...
  $\boldsymbol{v}_S(:, k) = e^{i\frac{2\pi f(k)}{c}\frac{d}{2}\cos(\varphi(j)-\boldsymbol{\phi_n})}$
  $\boldsymbol{V} = [\boldsymbol{v}_1, \boldsymbol{v}_2, ..., \boldsymbol{v}_S]$ ▷ Far-Field model.

  cvx_begin quiet ▷ Start cvx programming.
    variable $w(N_m)$ complex
      minimize $\|\boldsymbol{G}(:, :, k)\boldsymbol{w}(k) - \hat{\boldsymbol{B}}(:)\|_2^2$
    subject to
      $\boldsymbol{w}^H\boldsymbol{d} == 1$
      $\boldsymbol{w}^H\boldsymbol{w} <= 1/\gamma$
      $\boldsymbol{w}^H\boldsymbol{V} == \boldsymbol{0}$
  cvx_end ▷ End cvx programming.
end for

4.7 Generalized Sidelobe Canceller with Adaptive Blocking Matrix

4.7.1 The Algorithm

Adaptive beamformers are generally based on LMS[24] or NLMS[25] algorithms to minimize the output noise [7], and to extract a desired signal from many kinds of interfering signals, e.g., reverb, competing sources, etc. A combination of a fixed beamformer and adaptive algorithms yield the generalized sidelobe canceller. An efficient implementation of a GSC in frequency domain entails many computational savings [23]. The implementation used in this work consists of three blocks:

- the fixed beamformer,
- the adaptive blocking matrix, and
- the adaptive interference canceller.

The Fixed Beamformer

The fixed beamformer provides the reference signal for the adaptive interference canceller (AIC). Usually, a DS-BF enhances the components of the desired signal. An alternative is using a super-directive beamformer, e.g., the RLSFI-BF, or a combination of beamformers: a DS-BF for lower frequencies and a Dolph-Chebyshev design for higher frequencies.

The Adaptive Blocking Matrix

The adaptive blocking matrix is based on coefficient (un-)constrained adaptive filters which subtract the desired signal from the side lobe canceling path adaptively to prevent the cancellation of the desired signal by the AIC. Coefficient constraints restrict the GSC to cancel only a certain region around the desired signal.

The Adaptive Interference Cancellation

The adaptive interference canceller adaptively subtracts all signal components from the reference path—the path containing the signal enhanced by the fixed beamformer—which exhibit correlation between the adaptive interference canceller and the fixed beamformer output.

4.7.2 The Implementation

The implementation according to [23] requires time and frequency calculations as illustrated in Fig. 4.6. The use of a DFT-matrix $\boldsymbol{F}$ [24] simplifies the mathematical description of the algorithm,

$$\boldsymbol{F} = \frac{1}{\sqrt{2K}} \begin{pmatrix} 1 & 1 & 1 & \dots & 1 \\ 1 & w & w^2 & \dots & w^{2K-1} \\ 1 & w^2 & w^4 & \dots & w^{2(2K-1)} \\ \dots & \dots & \dots & \dots & \dots \\ 1 & w^{2K-1} & w^{2(2K-1))} & \dots & w^{(2K-1)(2K-1)} \end{pmatrix}$$

[24] Least Mean Square
[25] Normalized Least Mean Square

where $\boldsymbol{F}$ is a $(2K \times 2K)$-matrix, and $w = e^{-i\frac{2\pi}{2K}}$. The vector $\boldsymbol{y}_{FBF}$ contains a $(2K \times 1)$-set of output samples of the fixed beamformer and is defined as

$$\boldsymbol{y}_{FBF} = \begin{pmatrix} y_{FBF}(nK-K) \\ y_{FBF}(nK-K+1) \\ y_{FBF}(nK-K+2) \\ ... \\ y_{FBF}(nK) \\ ... \\ y_{FBF}(nK+K-1) \end{pmatrix}.$$

Consequently, the DFT of $\boldsymbol{y}_{FBF}$ is a $(2K \times 2K)$-diagonal-matrix

$$\boldsymbol{Y}_{FBF} = \text{diag}\left\{\boldsymbol{F} \cdot \boldsymbol{y}_{FBF}\right\}.$$

The Adaptive Blocking Matrix

The update-equation of the adaptive blocking matrix coefficients is

$$\boxed{\boldsymbol{B}_m(n+1) = \boldsymbol{B}_m(n) + \underbrace{\boldsymbol{G} \cdot \boldsymbol{\mu}(n) \cdot \boldsymbol{Y}_{FBF}^H(n) \cdot \boldsymbol{E}_{B,m}(n)}_{(2K\times 2K)\cdot(2K\times 2K)\cdot(2K\times 2K)\cdot(2K\times 1)}}, \tag{4.34}$$

where m is the microphone index and $1 \leq m \leq N$. The matrix

$$\boldsymbol{G} = \boldsymbol{F}\boldsymbol{g}\boldsymbol{F}^{-1}$$

constrains the gradient and ensures linear convolution [23] with $\boldsymbol{g} = \text{diag}\left\{(\boldsymbol{1}, \boldsymbol{0})\right\}$, where $\boldsymbol{0}$ and $\boldsymbol{1}$ are $(K \times 1)$-vectors. The $(2K \times 1)$-vector $\boldsymbol{B}_m(n)$ consists of the previous blocking coefficients, and $\boldsymbol{\mu}(n)$ describes the step-size according to

$$\boldsymbol{\mu}(n) = 2\mu\text{diag}\left\{\left(S^{-1}_{Y_{FBF},Y_{FBF}}(n,0), ..., S^{-1}_{Y_{FBF},Y_{FBF}}(n,2K-1)\right)\right\},$$

where μ is the fixed step-size parameter, and $S_{Y_{FBF},Y_{FBF}}(n,k)$ is the power estimate of the fixed-beamformer output of the k-th frequency bin with the forgetting factor λ

$$S_{Y_{FBF},Y_{FBF}}(n,k) = \lambda S_{Y_{FBF},Y_{FBF}}(n-1,k) + (1-\lambda)|Y_{FBF}(n,k)|^2$$

with $0 \leq k \leq 2K-1$ and $Y_{FBF}(n,k)$ as the magnitude of the k-th frequency bin of $\boldsymbol{Y}_{FBF}(n)$. The vector $\boldsymbol{E}_{B,m}$ is the result of the DFT of the error signal

$$\boldsymbol{E}_{B,m}(n) = \boldsymbol{F} \cdot \boldsymbol{e}_{B,m}(n), \tag{4.35}$$

$$\boldsymbol{e}_{B,m}(n) = \boldsymbol{x}_m(k-c) - \underbrace{\boldsymbol{v} \cdot \boldsymbol{F}^{-1} \cdot [\boldsymbol{Y}_{FBF}(n) \cdot \boldsymbol{B}_m(n)]}_{(2K\times 2K)\cdot(2K\times 2K)\cdot(2K\times 2K)\cdot(2K\times 1)},$$

where c describes the time lag because of block processing, $\boldsymbol{x}_m(n) = (\boldsymbol{0}, x_m(nK), ..., x_m(nK+K-1))^T$ is the input-data vector of channel m, and $\boldsymbol{v} = \text{diag}\left\{(\boldsymbol{0}, \boldsymbol{1})\right\}$ is a matrix which eliminates circular convolution effects, i.e. the first block or the first K samples are discarded and the second block is stored.

The Adaptive Interference Canceller

This section of the GSC requires the error signal of the ABM in frequency domain (4.35):

$$\boldsymbol{X}_{A,m}(n) = \text{diag}\left\{\boldsymbol{E}_{B,m}(n) + \boldsymbol{J} \cdot \boldsymbol{E}_{B,m}(n-1)\right\} \tag{4.36}$$

where $\boldsymbol{J} = \text{diag}\left\{(+1, -1, +1, -1, ..., -1)\right\}$ is a $(2L \times 2L)$-matrix and realizes a circular shift of L samples in frequency domain. This input matrix is fundamental for calculating the coefficients of the AIC according to

$$\boxed{\boldsymbol{A}_m(n+1) = \boldsymbol{A}_m(n) + \underbrace{\boldsymbol{G} \cdot \boldsymbol{\beta}(n) \cdot \boldsymbol{X}_{A,m}^H(n) \cdot \boldsymbol{E}_A(n)}_{(2K\times 2K)\cdot(2K\times 2K)\cdot(2K\times 2K)\cdot(2K\times 1)}}, \tag{4.37}$$

where

$$\boldsymbol{E}_A = \boldsymbol{F} \cdot (\boldsymbol{y}_{FBF}(n) - \boldsymbol{y}_{AIC}(n))$$

is the error of the AIC with

$$\boldsymbol{y}_{FBF}(n) = (\boldsymbol{0}, y_{FBF}(nK - c), ..., y_{FBF}(nK + K - 1 - c))^T,$$

and

$$\boldsymbol{\beta}(n) = 2\mu \text{diag}\left\{\left(S_{X_A,X_A}^{-1}(n,0), ..., S_{X_A,X_A}^{-1}(n, 2K-1)\right)\right\},$$

where μ is the fixed step-size parameter, and $S_{X_A,X_A}(n,k)$ is the power estimate of the signal defined in (4.36) output of the k-th frequency bin with the forgetting factor λ

$$S_{X_A,X_A}(n,k) = \lambda S_{X_A,X_A}(n-1,k) + (1-\lambda)\sum_{m=1}^{N} |X_{A,m}(n,k)|^2.$$

The output signal $\boldsymbol{y}_{AIC}(n)$—necessary for the computation of $\boldsymbol{E}_A(n)$—is the result of

$$\boldsymbol{y}_{AIC}(n) = \boldsymbol{F}^{-1}\left(\sum_{m=1}^{N} \boldsymbol{X}_{A,m}(n) \cdot \boldsymbol{A}_m(n)\right).$$

This work does not provide any pseudo code of the GSC because of its extensive source code. See framework function `beamdesignGSC.m` for more details.

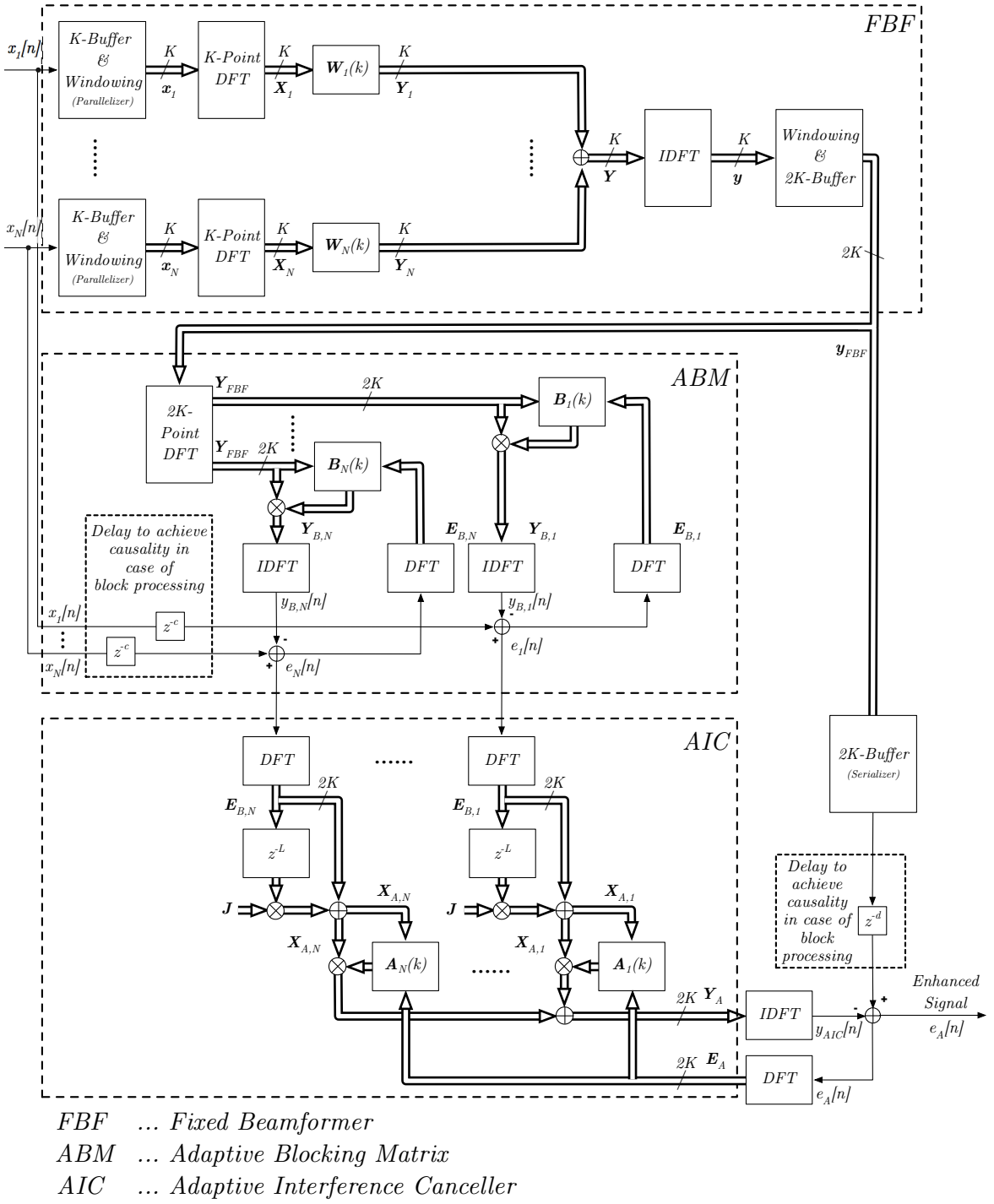


Figure 4.6: Block diagram of the robust GSC.

5 Beam-Pattern and Enhanced-Signal Measures

5.1 Measures for Beam Patterns

The following subsections contain information about the most common measures for beam pattern evaluation, i.e.

- the 3dB-Beamwidth (BW) [7],
- and the Main-to-Side-Lobe Ratio (MSR) [25].
- the Directivity Index (DI) [7][26],

5.1.1 Directivity Index (DI)

The beam patter is a complex function which represents a large amount of data. The directivity index reduces this huge amount to a small set of values or even a single value. According to [7] and [26] the directivity index for a given direction is the ratio of the received power from this direction to the received average power from all directions. It is a measure in dB and it represents the directivity of a microphone array. In general, it gives information about the noise suppression in case of isotropic ambient noise. The DI for a given frequency is

$$DI(f) = 10\log_{10} \frac{|U(f,\phi_s,\theta_s)|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}|U(f,\phi,\theta)|^2 d\theta d\phi} \tag{5.1}$$

and the total DI is

$$DI_{tot} = 10\log_{10} \int_0^{f_{max}} \frac{|U(f,\phi_s,\theta_s)|^2}{\frac{1}{4\pi}\int_0^{\pi}\int_0^{2\pi}|U(f,\phi,\theta)|^2 d\theta d\phi} df, \tag{5.2}$$

where $U(f,\phi,\theta)$ is the directivity of the array for a certain frequency f, azimuth ϕ, and elevation θ. The higher the directivity index the higher the increase in SNR in case of noisy listening situations, which results, e.g., in an improved speech recognition ability. The higher the DI the higher the ability in attenuating a competing speaker and the higher the increase in SNR in case of noisy listening situations which results, e.g., in an improved speech recognition ability.

5.1.2 3dB-Beamwidth (3dB-BW)

The 3dB-BW is the interval between two angles (in this work: ϕ_- and ϕ_+) where the main lobe of the beam pattern exhibits a gain of -3dB (half-power) and higher [7]. The smaller the beamwidth, the narrower the beam looking into the direction of the desired source, and the better the attenuation of competing sources next to the desired source.

5.1.3 Main-To-Side-Lobe Ratio (MSR)

The MSR is simply the gain-ratio between the peak of the main lobe and the peak of the side lobe which exhibits the lowest attenuation [25]. The higher the MSR, the better the attenuation of competing sources outside of the angular range of the main lobe. The MSR can be calculated as follows: set the first derivative of the beam pattern to zero to determine all extreme values, eliminate the extreme value at the steering direction ϕ_s, compute the second derivative to distinguish between minima and maxima, find the angle of the highest maxima, and determine the attenuation of this extreme value. The MSR is the difference between 0dB and the attenuation of this extreme value.

5.2 Measures for Enhanced Signals

The most reliable way to evaluate speech quality is the use of subjective listening tests. They are accurate, reliable, and repeatable; but also time-consuming and expensive. A more comfortable and favourable way is the use of objective measures, which highly correlate with subjective results. The only problem is to find such accurate and trustworthy objective measures. A human listener is able to evaluate, e.g., speech in different aspects:

- distortions that affect speech (speech distortion),
- distortions that affect the background noise (noise distortion).

It is impossible for a single objective measure to determine the influence of both types of distortions in a signal; the correlation between objective measures and subjective quality is only high for one type of distortion. Consequently, the use of composite measures based on various objective measures may exhibit a correlation of 95 % or higher [27]. Objective and composite measures may additionally depend on databases of subjective quality ratings. Common objective measures used for composite measures are

- segmental signal-to-noise ratio (segSNR) [28],
- weighted-slope spectral distance (WSS) [29],
- perceptual evaluation of speech quality (PESQ) [30],
- log-likelihood ratio (LLR) [31],
- Itakura-Saito distance (IS) [31],
- cepstrum distance (CEP) [32], and
- frequency-weighted segmental signal-to-noise ratio (fwsegSNR) [33]

to evaluate

- signal distortion,

- noise distortion, and
- overall quality.

Additionally, this work considers the Hidden Markov Model Speech Recognition Toolkit for word recognition as a measure of the quality of the enhanced signal. All measures assume a bandwidth of 8 kHz and a sampling frequency of $f_s = 16$ kHz. In this work, the measurements are performed by using the MATLAB source code for the PESQ implementation of [34]. The following sections cover descriptions of the used measures in this work.

5.2.1 Log-Likelihood Ratio (LLR)

The log-likelihood ratio is a common measure in speech research for comparing speech signals [35]. More precisely, it compares the fit of two different models. One of them is a special model—the null-model—based on the original speech signal, and the other one is the alternative model based on the enhanced signal. It is a LPC-based objective measure—it depends on the LPC of the enhanced signal—, and it computes the likelihood-ratio of both models [31]. The measure is defined as

$$LLR(\boldsymbol{a}_e, \boldsymbol{a}_o) = \ln \frac{\boldsymbol{a}_e^T \boldsymbol{R}_{oo} \boldsymbol{a}_e}{\boldsymbol{a}_o^T \boldsymbol{R}_{oo} \boldsymbol{a}_o},$$

where $\boldsymbol{a}_e$ is the LPC-vector of the enhanced signal, $\boldsymbol{a}_o$ is the LPC-vector of the original signal, and $\boldsymbol{R}_{oo}$ is the autocorrelation matrix of the original signal. The lower the LLR the better the speech quality of the enhanced signal.

5.2.2 Segmental Signal-To-Noise-Ratio (segSNR)

The segmental SNR is a time-domain measure. It computes the average of SNR values of a segment or frame of data [27]. It is defined as

$$segSNR = \frac{1}{K} \sum_{k=1}^{K} 10\log_{10} \left(\frac{\sum_{n=1}^{N} |s[n+kN]|^2}{\sum_{n=1}^{N} |\hat{s}[n+kN] - s[n+kN]|^2} \right),$$

where K is the number of frames which are part of a segment, N is the number of samples of a frame, $s[n]$ is the noise-free speech signal, and $\hat{s}[n]$ is the enhanced signal. The higher the measure the better the attenuation of noise and interferences during pauses.

5.2.3 Weighted-Slope Spectral Distance (WSS)

The weighted-slope spectral distance is a distance measure in frequency domain. It computes the weighted difference between the spectral slope of the original and the enhanced signal in each frequency band [27]. In general, it considers 25 critical bands [36]. It is computed as

$$WSS = \frac{1}{K} \sum_{k=1}^{K} \frac{\sum_{m=1}^{M} W(m,k)[S_o(m,k) - S_e(m,k)]^2}{\sum_{m=1}^{M} W(m,k)},$$

where M is the number of frequency bands, K is the number of frames which are part of a segment, $W(m,k)$ are the weights according to [29], $S_o(m,k)$ and $S_e(m,k)$ are the spectral slopes of the original and the enhanced signal for the j-th frequency band at frame m. The magnitude of each weight reflects whether the band is near a spectral peak or valley, and whether the peak

is the largest in the whole spectrum. The lower the WSS the better the speech quality of the enhanced signal.

5.2.4 Perceptual Evaluation of Speech Quality (PESQ)

The perceptual evaluation of speech quality measure is an objective measure and industry standard for objective speech quality evaluation, and generally used in telecommunications. The use of this measure avoids expensive and time-consuming listening tests, but it consumes a lot of computational resources. It analyzes speech signals sample-by-sample, and it has to be adapted before using it to exhibit a high correlation with the results of the subjective listening tests. Because of its huge complexity, this measure is not discussed in detail in this work. The higher the score, which lies between -0.5 and 4.5, the better the speech quality. See [27], [37], [38], and [39] for more details.

5.2.5 Composite Measures (C-SIG, C-BAK, C-OVRL)

The composite measures [27] evaluate signals filtered by speech enhancement algorithms. These measures relate to the objective measures mentioned above and subjective listening tests designed according to ITU-T recommendation P.835. The combination of the results of the objective measures and the listening tests yield three additional measures: a composite measure for signal distortion (C-SIG), a composite measure for background noise distortion (C-BAK), and a composite measure for overall speech quality (C-OVRL). These measures highly correlate with subjective ratings. According to [27] all three measures exhibit a correlation with subjective listening tests of 0.90 to 0.91 with signal distortion (C-SIG) and overall quality (C-ORVL). The higher the values the better the speech quality. The values for all measures are obtained by combining the LLR, WSS, PESQ, and segSNR linearly as quoted in [27].

The scale for C-SIG is as follows:

C-SIG Scale:	
Rating	Description
5	very natural, no degradation
4	fairly natural, little degradation
3	somewhat natural, somewhat degraded
2	fairly unnatural, fairly degraded
1	very unnatural, very degraded

The scale for C-BAK is as follows:

C-BAK Scale:	
Rating	Description
5	not noticeable
4	somewhat noticeable
3	noticeable but not intrusive
2	fairly conspicuous, somewhat intrusive
1	very conspicuous, very intrusive

The scale for C-ORVL is as follows:

C-OVRL Scale:	
Rating	Description
5	excellent
4	good
3	fair
2	poor
1	bad

5.2.6 Hidden Markov Model Speech Recognition Toolkit (HTK)

The Cambridge HTK [40] is a toolkit for building Hidden Markov Models in speech processing and is built according to the recipe of Keith Vertanen [41], which is available online[26]. It extracts equally spaced and discrete parameter vectors out of speech signals. The sequences of parameter vectors represent the speech waveform in terms of this parameters, i.e. it is possible to generate the original speech signal out of this parameters. One important task of this toolkit is to determine symbols, i.e. words or letters, out of the extracted parameter vectors. In this work, the toolkit performs speech recognition based on training- and evaluation processes, a simple grammar:

$$\langle\text{verb}\rangle\langle\text{colour}\rangle\langle\text{preposition}\rangle\langle\text{letter}\rangle\langle\text{digit}\rangle\langle\text{coda}\rangle,$$

and a triphone[27] dictionary. The training is necessary to tune the word recognizer with speaker-specific data from the CHiME-corpus to reduce the word error rate.

The word recognizer employs phonetic models based on the Wall-Street-Journal–Corpus (WSJ-Corpus) and the Acoustic Phonetic Continuous Speech Corpus (TIMIT-Corpus) provided by the SPSC and Stefan Petrik (Synvo[28]). Synvo also provided a Voiced-Unvoiced–Detector for the voice activity detection.

[26] HTK-Receipt: `http://www.keithv.com/software/htk/`
[27] Triphone is the abbreviation of three phonemes.
[28] Synvo: `http://www.synvo.com/`

6 Recording and Processing Environment

6.1 Recording Environment

The recording environment—the cocktail party room—at the Signal Processing and Speech Communication Laboratory Graz is a small conference room. Its details, e.g., the dimensions and the furniture, are shown in Fig. 6.1. The symbol representing a simplified loudspeaker unveils information about the source direction and the source distance relative to the UCA.

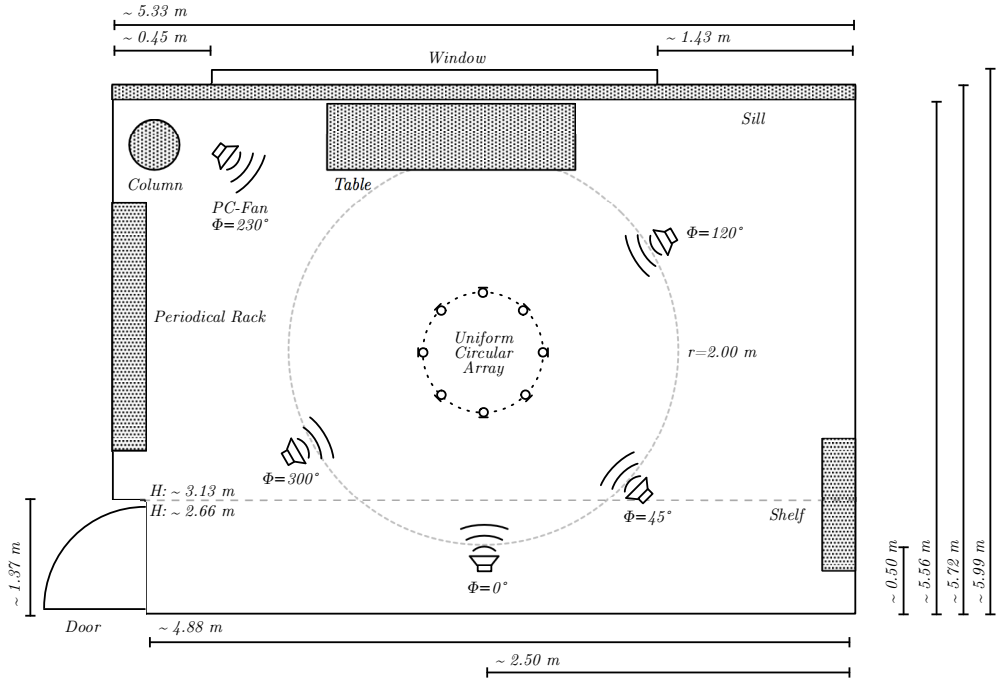


Figure 6.1: This figure shows the recording environment—the cocktail party room—at the Signal Processing and Speech Communication Laboratory Graz.

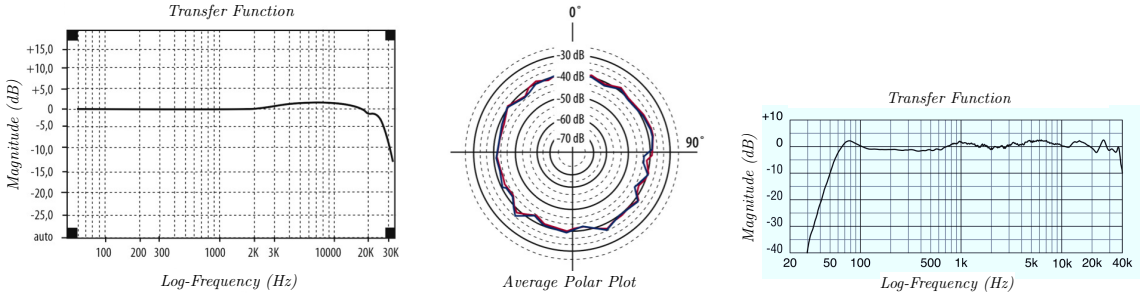
6.2 Recording Equipment

6.2.1 Loudspeakers

The Yamaha MSP5 Studio Loudspeaker, designed for serious monitoring, exhibits a flat frequency response up to 40 kHz and a uniform high frequency dispersion over 120 degrees. The frequency response of the loudspeaker is shown in Fig. 6.2(b).

6.2.2 Microphones

The Behringer Measurement Microphones ECM8000 are used for measuring the room and channel impulse response and for the recording. It is a precise electret condenser measurement microphone; it exhibits an ultra-linear frequency response and a well-balanced, true omnidirectional pattern. The deviations in the frequency response and the omnidirectional pattern are shown in Fig. 6.2(a).



(a) The microphone characteristics.

(b) The loudspeaker f-response.

Figure 6.2: This figure shows the the microphone characteristics and the loudspeaker frequency response.

6.2.3 Microphone Array

The microphone array shown in Fig. 6.4 consists of a wooden octagon with accurate and equidistant drillings on circles with different diameters, e.g., 20, 30, 40, and 55 cm, and drillings along a line which enables the use of ULAs. There are 24 drillings on a circle. The array is fixed on a metallic plate which is mounted on a stable microphone stand.

6.2.4 Sound Calibrator

The Sound Calibrator Cirrus Research PLC (CR: 511E, Serial: 039594, Class: 1L) is a portable, robust, and handy sound source for calibration of sound level meters. It exhibits a highly stable sound pressure level (94 ± 0.3 dB SPL) and calibration frequency (1000 ± 15 Hz). The calibrator is compatible with the Behringer Measurement Microphone ECM8000.

6.2.5 ADC and Audio Interface

The Behringer Ultragain Pro-8 Digital ADA8000 is a ultra high-quality 8-channel and high-end 24-bit A/D and D/A converter with a sampling rate of 44.1 and 48 kHz for digital recording, which includes eight new state-of-the-art microphone preamplifiers. It exhibits a 130 dB dynamic

range, a bandwidth ranging from 10 Hz until 200 kHz, and extremely low-noise and distortionless circuits.

The SM Pro Audio SM PR8E is a multi-channel preamplifier system for studio applications. It exhibits eight independent preamplifiers, a flat frequency response between 20 and 20 000 Hz—the absolute deviations are between 0 and 0.5 dB—, and high-quality components, which yield a high quality in audio processing.

The RME Fireface 800 is a 24-bit eight-channel FireWire® audio interfaces with a sampling frequency of 96 kHz.

6.3 Recording

6.3.1 Setup

The recording took place in a confined space: the cocktail party room with a closed window and a closed door. The heating system regulated the temperature to exactly 20.6°. The array with its 24 microphones—only two different diameters were considered (d=0.20m and d=0.55m, see Fig. 6.4 (a) and (d))—was placed in the middle of the room, and the loudspeakers were placed around the array at 0°, 45°, 120°, and 300° at a distance of 2 m relative to the center of the array. The loudspeaker in the upper-left part of Fig. 6.1 represents an interfering source in the form of a PC-fan. The UCA features a height of 1.175 m, and the center of the loudspeaker exhibits a height of 1.30 m (see Fig. 6.4 (b) and (c)). Section 6.2 lists the equipment used in this setup. Summarized, the recording required the following equipment:

Equipment:	
Array	01 x Wooden octagon plate
Microphones	24 x Behringer ECM8000
Loudspeakers	02 x Yamaha MSP5 Studio Loudspeakers
Calibrator	01 x Sound Calibrator Cirrus Research PLC (CR: 511E)
Audio-Interface	01 x RME Fireface 800
ADC/DAC	02 x Behringer Ultragain Pro-8 Digital ADA8000 01 x SM Pro Audio SM PR8E
PC	01 x Windows Notebook

6.3.2 Calibration

Each microphone including its corresponding channel was calibrated with the Cirrus CR: 511E. The calibrator, fixed on a microphone, generated a 1000 Hz oscillation with a SPL of 94 dB, which was recorded for 15 seconds. The subsequent computation of the RMS value of the recorded signal considered the signal between 4 and 12 seconds only. The computation of the compensation-gains α_n is as follows:

$$x_{RMS}^{(n)} = \sqrt{\frac{1}{M}\sum_{m}^{m+M} x^{(n)}[m]^2}$$
$$P^{(n)} = x_{RMS}^{(n)} \cdot x_{RMS}^{(n)}$$
$$P_{dB}^{(n)} = 10 \cdot \log_{10} P^{(n)}$$

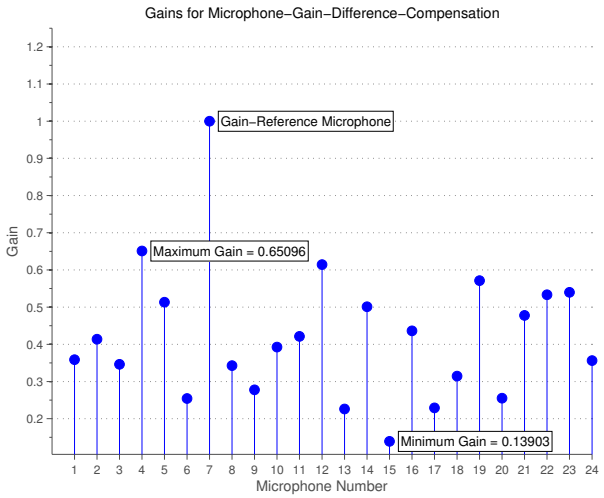


Figure 6.3: This figure shows the compensation-gains for each channel. The compensation gains scale each channel so that each recorded signal exhibits the same RMS-values.

where n is the channel number, $x[n]$ represents the recorded signal, and P is the power of the recorded signal. The channel, which exhibits the highest power $P_{ref,dB}$, is the reference channel for the computation of the compensation-gains α_n:

$$\Delta P_{dB}^{(n)} = P_{dB}^{(n)} - P_{ref,dB}$$
$$\alpha_n = 10^{\frac{\Delta P_{dB}^{(n)}}{10}}$$

The compensation gains α_n (see Fig. 6.3) scale each channel so that each recorded signal exhibits the same RMS-values:

$$x_{new}^{(n)}[n] = \frac{1}{\sqrt{\alpha_n}} \cdot x^{(n)}[n]$$

The reference channel is not affected. During the 30-minute calibration the air temperature remained constant at 20.6°.

The loudspeakers were calibrated relatively to each other. A microphone—the same for both loudspeakers—recorded a test signal 20 cm in front of the center of the loudspeaker membrane. Again, the power differences yielded the compensation-gains.

6.3.3 Test Signals

The test signals consist of twenty sentences, where each sentence exhibits the following grammar[29]:

$$\langle \text{verb} \rangle \langle \text{colour} \rangle \langle \text{preposition} \rangle \langle \text{letter} \rangle \langle \text{digit} \rangle \langle \text{coda} \rangle,$$

with

[29] CHiME-Corpus: `http://spandh.dcs.shef.ac.uk/projects/chime/research_corpus.html`

⟨verb⟩: {bin|lay|place|set},
⟨colour⟩: {blue|green|red|white},
⟨preposition⟩: {at|by|in|with},
⟨letter⟩: {a|b|c|...|x|y|z},
⟨digit⟩: {zero|one|two|...|seven|eight|nine}, and
⟨coda⟩: {again|now|please|soon}.

One possible sentence is, for instance, as follows:

Bin green with j four now.

There is a short break of two seconds between each sentence which lasts three seconds. The recorded signal consists of two different signals from two different directions and two different speakers coloured by the channel impulse responses which covers the whole transmission path, e.g., room impulse response, loudspeaker impulse response, etc. Both speakers use different words while they speak simultaneously, e.g.,

Speaker1: *Bin green in j four now.*
Speaker2: *Set white at w five soon.*

6.3.4 Playback and Recording

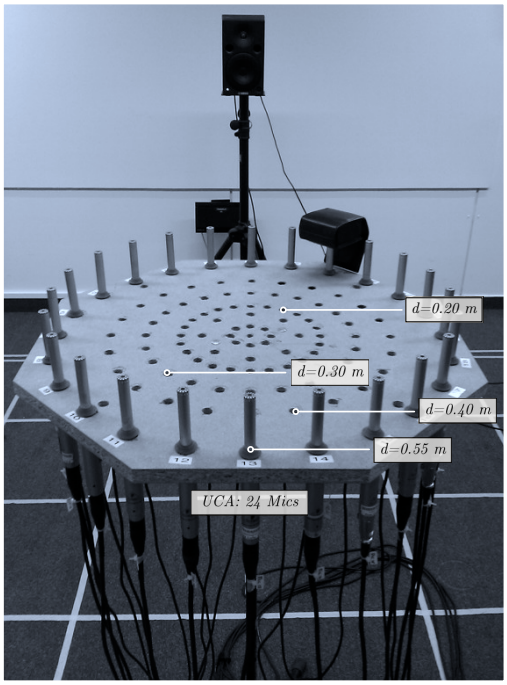
MATLAB 7.12.0 (R2011a)[30] generated the sentences automatically, and PureData[31] handled the playback, the proper timing of the signals, and the recording and storing on hard disk simultaneously.

6.4 Global Processing and Recording Parameters

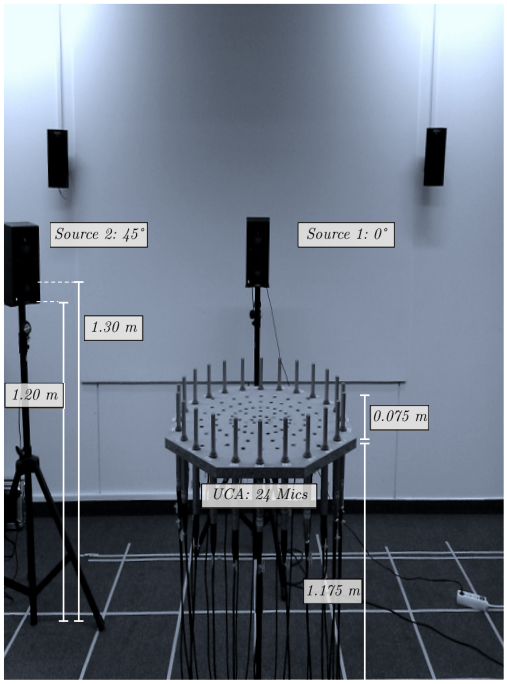
The following table contains information about global processing parameters in MATLAB and the recording parameters:

Frame Size:	2^8 samples
Window-Type:	sine-window
Overlap:	OLA[32] with 50 % overlapping
Sound Velocity:	343.57 m/s
Sampling Frequency:	48000 Hz
LP-Filter Cutoff Frequency:	16000 Hz
Angular Resolution:	01°
Source Elevation:	90°
ADC/DAC:	ideal
Microphone Type:	ideal omnidirectional
Assumed Sound Field for BFs	ideal and lossless far-field to avoid requirement of source distance

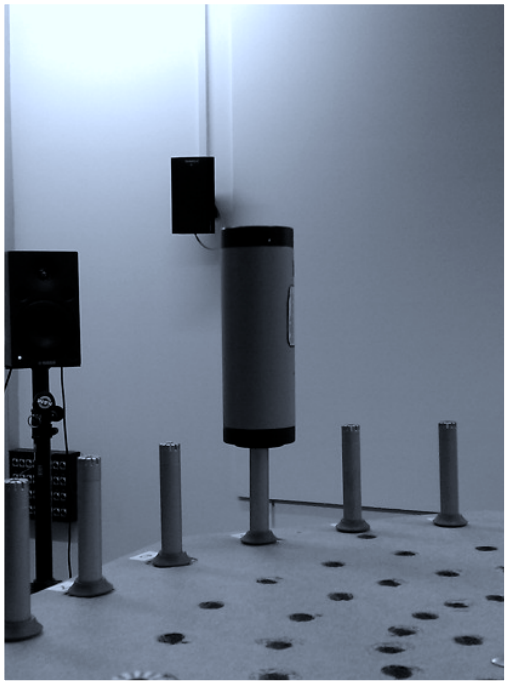
[30] MATLAB: `http://www.mathworks.de/products/matlab/index.html`
[31] PD: `http://puredata.info/`
[32] OLA - Overlap and Add



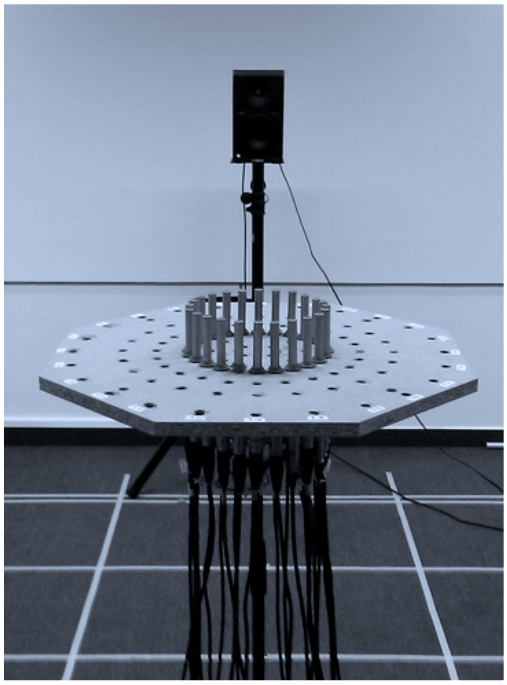
(a) Microphone array (d=0.55m).



(b) Microphone array and loudspeakers.



(c) Calibrator fixed on a microphone.



(d) Microphone array (d=0.20m).

Figure 6.4: This figure shows the microphone array with a diameter of d=0.55m (a,b,c) and d=0.20m (d), the loudspeakers (a-b), and the calibrator (c) used in this work.

7 Results

7.1 Beam Pattern Evaluation

The following measures unveil information about the beam pattern quality between 100 Hz and 16000 Hz:

- the 3dB-Beamwidth (3dB-BW),
- the Main-to-Side-Lobe Ratio (MSR), and
- the Directivity Index (DI).

This section covers the beam pattern evaluation of the data-independent beamformers only, because data-dependent beamformers change its beam pattern and measures after each time frame, which requires a huge number of additional pages and is therefore omitted in this work. The data-independent beamformers are

- the Delay&Sum-Beamformer (DS-BF)
- the Robust Least Squares Frequency Invariant Beamformer (RLSFI-BF), and
- the Multiple Null Synthesis RLSFI Beamformer (MNS-RLSFI-BF).

Animations show the changing measures and beam patterns over time in detail, which are available on demand.

3dB-Beamwidth

In case of a double-talk scenario the 3dB-BW of the main lobe has to be as small as possible to distinguish between two speaker sitting or standing close together. The higher the diameter of the UCA, the smaller the 3dB-BW as shown in Fig. A.1. A decomposition of the UCA into 2-element ULAs and a close look at the Dirichlet-Kernels $H(\eta)$ show that an increase in the microphone spacing d results in a decrease in beamwidth, because the null is reached earlier than with a lower spacing. An increase in the number of microphones leads to smaller, but hardly notable 3dB-BW values.

The DS-BF exhibits the highest 3dB-BW for lower frequencies before 2000 Hz (d=0.20m, mics=08), 3000 Hz (d=0.20m, mics=12), and 4000 Hz (d=0.20m, mics=24), and the lowest for higher frequencies. According to Fig. A.1 the 3dB-BW of the DS-BF exhibits the same progression for a constant diameter but for a different number of microphones.

In comparison to the DS-BF, both, the RLSFI- and the MNS-RLSFI-BF, exhibit a smaller 3dB-BW for lower frequencies and almost the same 3dB-BW—there are only small deviations—at higher frequencies.

Main-to-Side-Lobe Ratio

The higher the MSR, the higher the attenuation over all angles outside of the main lobe. According to Fig. A.2, the larger the diameter of the UCA the smaller the MSR over all frequencies, and the higher the number of microphones the higher the MSR.

The DS-BF does not exhibit a MSR at very low frequencies because of a missing side lobe—there is only a main lobe. That implies a higher spatial aliasing frequency in comparison to the RLSFI- and the MNS-RLSFI-BF. The DS-BF features high MSRs above 12000 Hz and low MSRs below that frequency. In comparison to the other data-independent beamformers the DS-BF exhibits a poorer MSR-progression the higher the number of microphones.

The RLSFI-BF features the best MSR-progressions. The smaller the number of microphones the better the MSR in comparison to the DS- and the MNS-RLSFI-BF. In general, its progressions are similar to the progressions of the MNS-RLSFI-BF for frequencies above 1000 Hz and identical for frequencies below that frequency, because the MNS-RLSFI-BF is a hybrid model which features a RLSFI-BF for lower frequencies.

Directivity Index

The higher the DI the higher the ability in attenuating a competing speaker and the higher the increase in SNR in case of noisy listening situations which results, e.g., in an improved speech recognition ability. According to Fig. A.3, an increase in the number of microphones leads to higher DI-values, whereas an increase in diameter yields a smaller frequency range which exhibits the highest DI-values.

The DS-BF exhibits good DIs for a low number of microphones. It approaches the indices of the other beamformers at higher frequencies. The higher the number of microphones the faster the approach. In comparison to the other beamformers the DS-BF features the smallest DIs at lower frequencies.

The RLSFI-BF exhibits the best overall directivity-index-progression. The MNS-RLSFI-BF features similar progressions for a high number of microphones, but the worst DIs for a low number of microphones due to the bad performance of the optimization algorithms for a lower number of microphones.

2D Beam Pattern

Fig. A.4 and A.5 show the beam patterns for different beamformers, numbers of microphones, and diameters, which reflect the results mentioned before. In general, one can see that an increase in diameter yields a smaller main lobe and smaller side lobes, a lower spatial aliasing frequency and a lower attenuation over all angles and frequencies. An increase in microphones results in a later decrease of attenuation of the side lobes and a higher attenuation over all angles

and frequencies.

In comparison to the DS-BF the RLSFI-BF features side lobes with higher attenuation at frequencies until 8000 Hz. The MNS-RLSFI-BF exhibits a beam pattern similar to the RLSFI-BF, but with an additional area exhibiting a very high attenuation due to the null-placement. The higher the number of microphones the more angles are affected from this area. A low number of microphones and a large diameter causes highly-attenuated frequency bands (see Fig. A.5e) due to the choice of parameters used for the optimization and a lower number of microphones.

3D Beam Pattern

The three dimensional beam pattern shown in Fig. A.6 and A.7—generally ignored in scientific papers—becomes attractive in case of reverberant environments, because it enables an estimation of the influence of reflections from different elevation angles, e.g., reflections from the floor or the ceiling, which may facilitate the decision to place a projector above or beside the array. In Fig. A.6a the DS-BF exhibits a widespread maine-lobe whereas the MNS-RLSFI-BF in Fig. A.6e features a flat main lobe at different elevation angles, which results in a higher attenuation of reflections impinging from different elevation angles. In case of positioning a projector above a UCA, the MNS-RLSFI-BF is a good choice because it exhibits a higher attenuation of signals with frequency $f = 2000$ Hz from $\theta = 0°$ than the DS-BF (compare Fig. A.6a and e). One way to improve the behaviour of the RLSFI- and the MNS-RLSFI-BF is to introduce additional constraints for different elevation angles.

7.2 Enhanced Signal Evaluation

The following tables summarize the numerical evaluations of the single- and double-talk scenarios based on synthetic and real data shown in Chapter B (Appendix). The first part of each table—all parts are separated by a double-line—unveils information about the best beamformers for each speaker and for a certain scenario (single- or double-talk). The second part summarizes the results of the beamformers' word recognition rates. The third part unveils information about the correlation of the objective measures and the word recognition rate of the beamformers, and the correlation of the objective measures and the audible quality.

The tables in Chapter B (Appendix) contain detailed information about the array properties and objective measures of filtered[33] and unfiltered[34] recordings. These tables consist of ten columns. The first one contains the number of microphones and the array diameter in centimeter, the second one contains the percentage of recognized words of the enhanced file (abbr.: WRe), the third one contains the percentage of recognized words of the nearest-microphone recording (abbr.: WRn), the sixth one contains log-likelihood ratio (abbr.: LLR), the seventh one contains the segmental SNR (abbr.: sSNR), the eighth one contains the Weighted-Slope Spectral Distance (abbr.: WSS), the ninth one contains the Perceptual Evaluation of Speech Quality (abbr.: PESQ) of the enhanced file, the tenth one contains the Composite Measure for Signal Distortion (abbr.: Csig) of the enhanced file, the eleventh one contains the Composite Measure for Background Noise Distortion (abbr.: Cbak) of the enhanced file, and the twelfth one contains the Composite Measure for Overall Speech Quality (abbr.: Covrl).

[33] WRe, LLR, sSNR, WSS, PESQ, Csig, Cbak, Covrl
[34] WRn

The two remaining measures appearing in these tables—the improvement in the global signal to interference plus noise ratio (iGSINR) and a gain measure (GAIN)—are not discussed in this work because of their low significance in the evaluation.

Synthetic Data

The synthetic scenarios,

- 1st Scenario: Double-Talk (Speaker 03[35] and 12 @ MRA = $\{0°, 45°\}$) and
- 2nd Scenario: Double-Talk (Speaker 03 and 12 @ MRA = $\{300°, 120°\}$),

exhibit an ideal array aperture, no deviations in microphone positions, no mismatches in microphones and loudspeakers, no room impulse responses, and a constant and well-known sound velocity $c = 343$ m/s. The synthetic audio signals match the audio signals played-back by the loudspeakers in the real scenarios, except that they features ideal time-shifts and attenuations between the loudspeakers and the microphones and no influence of the loudspeaker characteristics.

In the 1st scenario both speakers are close together. The numerical results match the graphical evaluation in Section 7.1 and the data-independent and data-dependent beamformers' theoretical behaviour mentioned in Chapter 4. The MNS-RLSFI-BF exhibits the highest word recognition rate of the data-independent (Sp.1 = 86.66 % and Sp.2 = 81.66 %), the GSC of the data-dependent beamformers (Sp.1 = 83.33 % and Sp.2 = 79.17 %). Moreover, both exhibit the highest difference between the word recognition rate of the enhanced signal and the word recognition rate of the signal captured by the nearest microphone: the MNS-RLSFI-BF achieves Sp.1 = 36.66 % and Sp.2 = 35.84 %, the GSC Sp.1 = 33.33 % and Sp.2 = 31.67 %. The data-independent beamformers exhibit a better performance than the data-dependent beamformers. According to the word recognition rates the MNS-RLSFI-BF is better than the RLSFI-BF and the DS-BF (in this order) which corresponds to the theory. Among the word recognition rates, the MNS-RLSFI-BF always achieves the highest Cbak for both speakers, but it does not achieve the highest PESQ and Corvl in both cases due to a different pronunciation of the words. The highest PESQ is achieved with an array diameter of d=0.55 m, the highest Csig with d=0.20 m, the highest Cbak with d=0.20 m, and the highest Corvl with d=0.20 m. Both, the word recognition rates and the other objective measures, exhibit a high correlation. See Tab. 7.1 for more details.

In the 2nd scenario both speakers are talking face-to-face. Again, the numerical results match the graphical evaluation in Section 7.1 and the data-independent and data-dependent beamformers' theoretical behaviour mentioned in Chapter 4. The MNS-RLSFI-BF exhibits the highest word recognition rate of the data-independent (Sp.1 = 89.17 % and Sp.2 = 91.66 %), the GSC of the data-dependent beamformers (Sp.1 = 78.33 % and Sp.2 = 71.67 %). The MNS-RLSFI-BF achieves the highest difference between the word recognition rate of the enhanced signal and the word recognition rate of the signal captured by the nearest microphone: the MNS-RLSFI-BF achieves Sp.1 = 37.50 % and Sp.2 = 45.00 %, the GSC Sp.1 = 27.50 % and Sp.2 = 25.84 %. The data-independent beamformers exhibit a better performance than the data-dependent beamformers. According to the word recognition rates the MNS-RLSFI-BF is better than the RLSFI-BF and the DS-BF, but both, the RLSFI- and the DS-BF exhibit the same performance due to the positions of both speaker: a smaller beamwidth—the main advantage of the RLSFI-BF—does not matter, if both speakers are talking face-to-face. Among the word recognition

[35] The number of the speaker corresponds to the number of the speaker in the CHiME-database.

rates, the MNS-RLSFI-BF achieves the highest PESQ and composite measures. The highest PESQ is achieved with an array diameter of d=0.55 m, the highest Csig with d=0.55 m, the highest Cbak with d=0.20 m, and the highest Corvl with d=0.55 m. Both, the word recognition rates and the other objective measures, exhibit a high correlation in case of data-independent and data-dependent beamformers. See Tab. 7.2 for more details.

Summarizing and comparing the results of both double-talk scenarios with synthetic data yield the following hypothesis:

- the best data-independent beamformer is the MNS-RLSFI-BF,
- the best data-dependent beamformer is the GSC,
- the numerical results match the beamformers' theoretical behaviour,
- medium correlation between the word recognition rate and all other obj. measures for DI-BF[36] and side-by-side scenarios,
- high correlation between the word recognition rate and all other obj. measures for DI-BF[37] and face-to-face scenarios,
- medium correlation between the word recognition rate and all other obj. measures for DD-BF[38] and side-by-side scenarios,
- high correlation between the word recognition rate and all other obj. measures for DD-BF[39] and face-to-face scenarios,
- a small diameter yields the best perceptual results when both speakers are close together,
- a large diameter yields the best perceptual results when both speakers talk face-to-face.

[36] DI-BF - Data Independent Beamformer
[37] DI-BF - Data Independent Beamformer
[38] DD-BF - Data Dependent Beamformer
[39] DD-BF - Data Dependent Beamformer

1st Scenario: Double-Talk (Speaker 03 and 12 @ MRA = $\{0°, 45°\}$)

	Speaker 1 ($\phi_s = 00°$)						**Speaker 2** ($\phi_s = 45°$)					
	DI			**DD**			**DI**			**DD**		
	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC
Best BF			x			x			x	x		
Setups exh. Improvements	6	6	6	6	0	6	6	6	6	6	0	6
Highest WRe-Rate	84.17 %	85.00 %	86.66 %	76.66 %	38.33 %	83.33 %	78.33 %	77.50 %	81.66 %	80.83 %	36.66 %	79.17 %
Setup	24/55	24/20	12/55	08/20	12/55	24/55	12/55	24/20	08/55	08/20	08/55	12/55
Highest ΔWRe-WRn	34.17 %	34.17 %	36.66 %	25.84 %	-11.65 %	33.33 %	31.66 %	30.83 %	35.84 %	34.16 %	-11.66 %	31.67 %
Setup	24/55	24/20	12/55	08/20	12/55	24/55	24/55	24/20	08/55	08/55	08/55	24/55
Highest PESQ		x				x			x	x		
Highest Csig	x					x		x				x
Highest Cbak			x			x			x			x
Highest Covrl			x	x				x				x

Table 7.1: This table summarizes the numerical results of the double-talk scenario 1 based on synthetic data in Chapter B (Appendix).

2nd Scenario: Double-Talk (Speaker 03 and 12 @ MRA = $\{300°, 120°\}$)

	Speaker 1 ($\phi_s = 300°$)						**Speaker 2** ($\phi_s = 120°$)					
	DI			**DD**			**DI**			**DD**		
	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC
Best BF			x			x			x			x
Setups exh. Improvements	6	6	6	0	0	6	6	6	6	0	0	6
Highest WRe-Rate	85.00 %	85.00 %	89.17 %	35.00 %	32.50 %	78.33 %	87.50 %	87.50 %	91.66 %	37.50 %	25.00 %	71.67 %
Setup	(24,55)	(24,55)	(24,55)	(12,55)	(12,55)	08/20	(24,55)	(24,55)	(08,20)	(24,55)	(08,55)	08/55
Highest ΔWRe-WRn	33.34 %	33.34 %	37.50 %	-17.50 %	-20.00 %	27.50 %	41.66 %	41.66 %	45.00 %	-08.33 %	-20.83 %	25.84 %
Setup	(08,20)	(08,20)	(08,55)	(12,55)	(12,55)	08/20	(24,55)	(24,55)	(12,55)	(24,55)	(08,55)	08/55
Highest PESQ			x			x			x			x
Highest Csig			x			x			x			x
Highest Cbak			x			x			x			x
Highest Covrl			x			x			x			x

Table 7.2: This table summarizes the numerical results of the double-talk scenario 2 based on synthetic data in Chapter B (Appendix).

Real Data (Single-Talk)

The real scenarios,

- 1st Scenario: Single-Talk (Speaker 03 and 12 @ MRA = $\{0°, 45°\}$) and
- 2nd Scenario: Single-Talk (Speaker 03 and 12 @ MRA = $\{300°, 120°\}$),

exhibit a non-ideal array aperture, small deviations in microphone positions, small mismatches in microphones and loudspeakers—both are calibrated—, reverb, and a varying but known sound velocity $c = 343 \pm 0.5$ m/s. Both speakers talk separately in the following two scenarios. This evaluation is necessary for the determination of the highest achievable values.

In the 1st scenario the numerical results match the graphical evaluation in Section 7.1 and the data-independent beamformers' theoretical behaviour mentioned in Chapter 4, but the numerical results of the data-dependent beamformers do not. The most ordinary data-dependent beamformer, the MPDRDL-BF, exhibits the best results. All data-independent beamformers, the DS-, the RLSFI-, and the MNS-RLSFI-BF, exhibit the highest word recognition rate of the data-independent (Sp.1 = 95.83 % and Sp.2 = 89.17 %), the MPDRDL-BF of the data-dependent beamformers (Sp.1 = 95.83 % and Sp.2 = 83.33 %). Moreover, all four beamformers exhibit the highest difference between the word recognition rate of the enhanced signal and the word recognition rate of the signal captured by the nearest microphone: the DS-, the RLSFI-, and the MNS-RLSFI-BF achieve Sp.1 = 04.16 % and Sp.2 = 10.00 %, the MPDRDL-BF Sp.1 = 04.16 % and Sp.2 = 03.33 %. The data-independent beamformers exhibit a better performance than the data-dependent beamformers. The MNS-RLSFI-BF is as good as the RLSFI-BF and the DS-BF. Among the word recognition rates, all data-independent beamformers achieve the highest objective measures for both speakers, the MPDRDL-BF achieves the highest PESQ, Cbak, and Corvl in both cases. The highest PESQ is achieved with an array diameter of d=0.20 m, the highest Csig with d=0.20 m, the highest Cbak with d=0.20 m, and the highest Corvl with d=0.20 m. Both, the word recognition rates and the other objective measures, exhibit a high correlation in case of data-independent, and a low correlation in case of data-dependent beamformers. See Tab. 7.3 for more details.

The 2nd scenario exhibits the same tendencies as the first one. The numerical results match the graphical evaluation in Section 7.1 and the data-independent beamformers' theoretical behaviour mentioned in Chapter 4, but the numerical results of the data-dependent beamformers do not. The most ordinary beamformer, the MPDRDL-BF, exhibits the best results. All data-independent beamformers, the DS-, the RLSFI-, and the MNS-RLSFI-BF, exhibit the highest word recognition rate of the data-independent (Sp.1 = 95.83 % and Sp.2 = 88.33 %), the MPDRDL-BF of the data-dependent beamformers (Sp.1 = 93.33 % and Sp.2 = 83.33 %). Moreover, all four beamformers exhibit the highest difference between the word recognition rate of the enhanced signal and the word recognition rate of the signal captured by the nearest microphone: the DS-, the RLSFI-, and the MNS-RLSFI-BF achieve Sp.1 = 03.33 % and Sp.2 = 09.16 %, the MPDRDL-BF Sp.1 = 00.83 % and Sp.2 = 04.16 %. The data-independent beamformers exhibit a better performance than the data-dependent beamformers. The MNS-RLSFI-BF is as good as the RLSFI-BF and the DS-BF. Among the word recognition rates, all data-independent beamformers achieve the highest objective measures for both speakers, the MPDRDL-BF does not achieve the highest PESQ, Cbak, and Corvl in both cases. The highest PESQ is achieved with an array diameter of d=0.20 m, the highest Csig with d=0.20 m, the highest Cbak with d=0.20 m, and the highest Corvl with d=0.20 m. Both, the word recognition rates and the other objective measures, exhibit a high correlation in case of data-independent, and no correlation in case of data-dependent beamformers. See Tab. 7.4 for more details.

Summarizing and comparing the results of both single-talk scenarios with real data yield the following hypothesis:

- each data-independent beamformer is suitable for the single-speaker scenario,
- the best data-dependent beamformer is the MPDRDL-BF,
- the numerical results match the beamformers' theoretical behaviour in case of DI-BF,
- high correlation between the word recognition rate and all other obj. measures for DI-BF for side-by-side and face-to-face scenarios,
- no correlation between the word recognition rate and all other obj. measures for DD-BF for side-by-side and face-to-face scenarios,
- a small diameter yields the best perceptual results when both speakers are close together,
- a small diameter yields the best perceptual results when both speakers talk face-to-face.

Real Data (Double-Talk)

The real scenarios,

- 1st Scenario: Double-Talk (Speaker 03 and 12 @ MRA = $\{0°, 45°\}$) and
- 2nd Scenario: Double-Talk (Speaker 03 and 12 @ MRA = $\{300°, 120°\}$),

exhibit a non-ideal array aperture, small deviations in microphone positions, small mismatches in microphones and loudspeakers—both are calibrated—, reverb, and a varying but known sound velocity $c = 343 \pm 0.5$ m/s. Both speaker talk simultaneously.

In the 1st scenario both speakers are close together. The DS-BF exhibits the highest word recognition rate of the data-independent (Sp.1 = 68.33 % and Sp.2 = 50.00 %), the MPDRDL-BF of the data-dependent beamformers (Sp.1 = 64.17 % and Sp.2 = 38.33 %). Moreover, the DS-BF and the GSC exhibit the highest difference between the word recognition rate of the enhanced signal and the word recognition rate of the signal captured by the nearest microphone: the DS-BF achieves Sp.1 = 10.83 % and Sp.2 = 16.67 %, the GSC Sp.1 = 09.17 % and Sp.2 = 04.17 %. The data-independent beamformers exhibit a better performance than the data-dependent beamformers. Among the word recognition rates, the DS-BF achieves the highest objective measures for both speakers, the MPDRDL-BF achieves the highest Cbak for one speaker only. The GSC exhibits the highest Covrl in both cases. The highest PESQ is achieved with an array diameter of d=0.20 m, the highest Csig with d=0.20 m, the highest Cbak with d=0.20 m, and the highest Corvl with d=0.20 m. Both, the word recognition rates and the other objective measures, exhibit a high correlation in case of data-independent, and a low correlation in case of data-dependent beamformers. See Tab. 7.5 for more details.

In the 2nd scenario both speakers are talking face-to-face. The DS-BF exhibits the highest word recognition rate of the data-independent (Sp.1 = 66.67 % and Sp.2 = 48.33 %), the GSC of the data-dependent beamformers (Sp.1 = 57.50 % and Sp.2 = 34.17 %). Moreover, the DS-BF and the GSC exhibit the highest difference between the word recognition rate of the enhanced signal and the word recognition rate of the signal captured by the nearest microphone: the DS-BF achieves Sp.1 = 15.84 % and Sp.2 = 13.34 %, the GSC Sp.1 = 06.67 % and Sp.2 = 03.34 %. The data-independent beamformers exhibit a better performance than the data-dependent beamformers. Among the word recognition rates, both, the DS-BF and the GSC, achieve the

highest objective measures in both directions. The highest PESQ is achieved with an array diameter of d=0.20 m, the highest Csig with d=0.20 m, the highest Cbak with d=0.20 m, and the highest Corvl with d=0.20 m. Both, the word recognition rates and the other objective measures, exhibit a high correlation in case of data-independent, and a high correlation in case of data-dependent beamformers. See Tab. 7.6 for more details.

Summarizing and comparing the results of both double-talk scenarios with real data yield the following hypothesis:

- the best data-independent beamformer is the DS-BF,
- the best data-dependent beamformer is the GSC,
- the numerical results match the beamformers' theoretical behaviour in case of DI-BF,
- high correlation between the word recognition rate and all other obj. measures for DI-BF for side-by-side and face-to-face scenarios ,
- high correlation between the word recognition rate and all other obj. measures for DD-BF for face-to-face scenarios,
- no correlation between the word recognition rate and all other obj. measures for DD-BF for side-by-side scenarios,
- a small diameter yields the best perceptual results when both speakers are close together,
- a small diameter yields the best perceptual results when both speakers talk face-to-face.

1st Scenario: Single-Talk (Speaker 03 and 12 @ MRA = $\{0°, 45°\}$)

	Speaker 1 ($\phi_s = 00°$)						**Speaker 2** ($\phi_s = 45°$)					
	DI			**DD**			**DI**			**DD**		
	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC
Best BF	x	x	x	x			x	x	x	x		
Setups exh. Improvements	6	6	6	3	2	0	6	6	6	5	0	0
Highest WRe-Rate	95.83 %	95.83 %	95.83 %	95.83 %	94.17 %	91.67 %	89.17 %	89.17 %	89.17 %	83.33 %	73.33 %	77.50 %
Setup	24/20	24/20	24/20	08/20	24/55	08/20	24/55	24/55	24/55	12/55	08/20	08/20
Highest ΔWRe-WRn	04.16 %	04.16 %	04.16 %	04.16 %	01.66 %	00.00 %	10.00 %	10.00 %	10.00 %	03.33 %	-04.17 %	00.00 %
Setup	24/20	24/20	24/20	08/20	24/20	08/20	24/55	24/55	24/55	12/20	08/20	08/20
Highest PESQ	x	x	x		x		x	x	x		x	
Highest Csig	x	x	x			x	x	x	x		x	
Highest Cbak	x	x	x		x		x	x	x		x	
Highest Covrl	x	x	x		x		x	x	x		x	

Table 7.3: This table summarizes the numerical results of the single-talk scenario 1 based on real data in Chapter B (Appendix).

1st Scenario: Double-Talk (Speaker 03 and 12 @ MRA = $\{0°, 45°\}$)

	Speaker 1 ($\phi_s = 300°$)						**Speaker 2** ($\phi_s = 120°$)					
	DI			**DD**			**DI**			**DD**		
	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC
Best BF	x			x			x			x		
Setups exh. Improvements	5	5	4	4	2	3	6	3	4	4	2	3
Highest WRe-Rate	68.33 %	64.17 %	62.50 %	64.17 %	56.67 %	64.17 %	50.00 %	41.67 %	40.00 %	38.33 %	25.83 %	37.50 %
Setup	24/55	08/55	12/20	08/55	24/20	24/20	24/55	08/55	24/20	12/55	12/55	08/55
Highest ΔWRe-WRn	10.83 %	07.50 %	07.50 %	06.67 %	01.67 %	09.17 %	16.67 %	08.34 %	06.67 %	07.50 %	-05.00 %	04.17 %
Setup	24/20	08/20	12/20	08/20	24/20	24/20	24/55	08/20	24/20	12/55	12/55	08/55
Highest PESQ	x					x	x				x	
Highest Csig	x						x					x
Highest Cbak	x			x		x	x				x	
Highest Covrl	x					x	x					x

Table 7.4: This table summarizes the numerical results of the double-talk scenario 1 based on real data in Chapter B (Appendix).

2nd Scenario: Single-Talk (Speaker 03 and 12 @ MRA = $\{300°, 120°\}$)

	Speaker 1 ($\phi_s = 300°$)						**Speaker 2** ($\phi_s = 120°$)					
	DI			**DD**			**DI**			**DD**		
	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC
Best BF	x	x	x	x			x	x	x	x		
Setups exh. Improvements	3	3	3	1	1	0	4	4	4	3	0	1
Highest WRe-Rate	95.83 %	95.83 %	95.83 %	93.33 %	92.50 %	89.17 %	88.33 %	88.33 %	88.33 %	83.33 %	74.17 %	75.83 %
Setup	24/20	24/20	24/20	08/55	08/20	12/55	24/55	24/55	24/55	24/55	08/20	08/20
Highest ΔWRe-WRn	03.33 %	03.33 %	03.33 %	00.83 %	00.83 %	-03.34 %	09.16 %	09.16 %	09.16 %	04.16 %	00.00 %	01.66 %
Setup	08/20	08/20	08/20	08/20	08/20	08/20	24/55	24/55	24/55	24/55	08/20	08/20
Highest PESQ	x	x	x			x	x	x	x		x	
Highest Csig	x	x	x			x	x	x	x		x	
Highest Cbak	x	x	x		x		x	x	x		x	
Highest Covrl	x	x	x			x	x	x	x		x	

Table 7.5: This table summarizes the numerical results of the single-talk scenario 2 based on real data in Chapter B (Appendix).

2nd Scenario: Double-Talk (Speaker 03 and 12 @ MRA = $\{300°, 120°\}$)

	Speaker 1 ($\phi_s = 300°$)						**Speaker 2** ($\phi_s = 120°$)					
	DI			**DD**			**DI**			**DD**		
	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC	DS	RLSFI	MNS	MPDRDL	MPDRVL	GSC
Best BF	x					x	x					x
Setups exh. Improvements	6	2	2	0	1	5	5	2	2	0	0	2
Highest WRe-Rate	66.67 %	61.67 %	61.67 %	48.33 %	52.50 %	57.50 %	49.17 %	36.67 %	35.83 %	24.17 %	23.33 %	34.17 %
Setup	24/55	08/20	08/20	12/55	24/55	08/20	24/55	08/55	08/55	08/20	12/55	08/20
Highest ΔWRe-WRn	15.84 %	10.84 %	10.84 %	-02.50 %	01.67 %	06.67 %	13.34 %	03.34 %	02.50 %	-06.66 %	-10.83 %	03.34 %
Setup	24/55	08/20	08/20	12/55	24/55	08/20	24/55	08/55	08/55	08/20	08/20	08/20
Highest PESQ	x					x	x					x
Highest Csig	x					x	x					x
Highest Cbak	x					x	x					x
Highest Covrl	x					x	x					x

Table 7.6: This table summarizes the numerical results of the double-talk scenario 2 based on real data in Chapter B (Appendix).

8 Conclusion and Future Work

8.1 Conclusion

In double-talk scenarios with simulated wave-propagation, free-field conditions, and perfect sound capturing without any deviations in microphone positions, loudspeaker characteristics, and microphone characteristics, the beam pattern and numerical evaluations confirm the data-independent beamformers' theoretical behaviour and highlight the advantages of the new established MNS-RLSFI-BF, which achieves the highest word recognition rates when both speakers are talking face-to-face or side-by-side. The differences between the word recognition rats of the DS-BF and the RLSFI-BF are smaller than the differences between the DS-BF and the MNS-RLSFI-BF or the RLSFI-BF and the MNS-RLSFI-BF, which underlines the high performance of the MNS-RLSFI-BF. A high correlation between the word recognition rates and the remaining objective measures—the PESQ, C-SIG, C-BAK, and C-OVRL—in case of face-to-face scenarios and a medium correlation in case of side-by-side scenarios show that a high speech quality of the enhanced signal corresponds to a high word recognition rate. The MNS-RLSFI-BF achieves the highest C-BAK in side-by-side and face-to-face scenarios and the highest PESQ and composite measures (C-SIG, C-BAK, and C-OVRL) in face-to-face scenarios.

In case of data-dependent beamformers the GSC yields the highest word recognition rates for side-by-side and face-to-face scenarios. It is noteworthy that the GSC is the most complex and CPU-intensive beamformer in this work. The MPDRVL-BF always exhibits lower word recognition rates than the nearest microphone for both double-talk scenarios, and so is the MPDRDL-BF for face-to-face scenarios. A medium and high correlation between the word recognition rates and the remaining objective measures in case of side-by-side and face-to-face scenarios show that a high speech quality of the enhanced signal corresponds to a high word recognition rate. The GSC achieves the highest C-SIG and C-BAK in side-by-side scenarios and the highest PESQ and composite measures in face-to-face scenarios.

Additionally, a small array diameter yields the best perceptual results when both speakers are close together, whereas a large diameter yields the best perceptual results when both speakers talk face-to-face.

In real double-talk scenarios with minimal deviations in temperature ($\pm 0.5^\circ$), small deviations in loudspeaker characteristics, microphone characteristics, and microphone positions (a non-ideal array aperture), the DS-BF always achieves the highest word recognition rates among the data-

independent beamformers due to the mismatches and deviations which affect the performance of the RLSFI-BF and the MNS-RLSFI-BF. The differences between the word recognition rates of all data-independent beamformers are larger than in case of the synthetic scenario which points out the robustness of the DS-BF and the performance loss of the RLSFI-BF and the MNS-RLSFI-BF due to the mismatches and deviations, which matches the super-directive beamformers' theoretical behaviour in case of real scenarios. There is a high correlation between the word recognition rates and the remaining objective measures in case of side-by-side and face-to-face scenarios. Thus, a high speech quality of the enhanced signal corresponds to a high word recognition rate. The DS-BF achieves the best PESQ and composite measures in side-by-side and face-to-face scenarios.

In case of data-dependent beamformers the MPDRDL-BF yields the highest word recognition rates in side-by-side, the GSC in face-to-face scenarios. There is a high correlation between the word recognition rates and the remaining objective measures in case of face-to-face scenarios, and no correlation in case of side-by-side scenarios. The GSC achieves the highest PESQ and composite measures in case of side-by-side and face-to-face scenarios.

Additionally, a small array diameter yields the best perceptual results when both speakers talk side-by-side or face-to-face.

What this all amounts to is that data-dependent beamformers exhibit a lower performance and a higher sensitivity to mismatches and deviations than super-directive data-independent beamformers. A comparison between the results of the synthetic and real double-talk scenarios show that reverb, mismatches, and deviations in microphone characteristics, loudspeaker characteristics, and microphone positions yield a decrease in the absolute word recognition rate between 10 % and 60 %.

8.2 Future Work

There are three major areas where modifications may lead to improvements in the reliability of the results and speech quality:

- the audio database,
- the microphone array, and
- the beamformer.

The CHiME-Database exhibits a few problems which can not be solved without changing its audio files. There are just a few records whose spoken sentences start and stop at the same instant of time; but this is a requirement for fair conditions in double-talk scenarios. In many cases one speaker starts or stops before or after the other one. Consequently, there is at least one word in each sentence without double-talk and, thus, always recognized successfully. A new database with male and female speakers which exhibits words only (no sentences) and the same RMS-values for all files would simplify the composition of different sentences with a specified grammar, e.g., the same grammar as used in the CHiME-Database. Moreover, a new plate with more accurate, intelligently placed drillings for different array geometries, e.g., randomly positioned microphones, and a more stable stand would eliminate the mismatches and deviations due to skewed microphones. Furthermore, the use of a genetic algorithm may improve beamformer specific parameters that may lead to higher word recognition rates and hearable improvements in speech quality. The use of intelligent hybrid- or Eigen-beamformers may improve the word recognition rates and speech quality too.

Bibliography

[1] X. Zhang and J. Hansen, "Csa-bf: A constrained switched adaptive beamformer for speech enhancement and recognition in real car environments," *Speech and Audio Processing, IEEE Transactions*, vol. 11, no. 6, pp. 733–745, November 2003.

[2] P. Vary and R. Martin, *Digital Speech Transmission*. John Wiley and Sons, Inc., March 2006.

[3] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing*, ser. Springer Topics in Signal Processing. Springer, January 2008, vol. 1.

[4] J. Benesty, Y. Huang, and M. Sondhi, *Springer Handbook of Speech Processing*. Springer, December 2007.

[5] F. Zotter and H. Pomberger, "Acoustic holography and holophony," Lecture Notes, Institute of Electronic Music and Acoustics, Inffeldgasse 10/3, 8010 Graz, Austria, October 2011.

[6] H. Teutsch, "Wavefield decomposition using microphone arrays and its application to acoustic scene analysis," Ph.D. dissertation, Technische Fakultät der Friedrich-Alexander Universität Erlangen-Nürnberg, Erlangen-Nürnberg, Germany, October 2005.

[7] I. Tashev, *Sound Capture and Processing: Practical Approaches*. John Wiley and Sons, Inc., July 2009.

[8] ——, "International conference for multimedia and expo icme 2004," Taipei, Taiwan, June 2004.

[9] M. Ehsan and G. Kubin, "Frame change ratio: A measure to model short-time stationarity of speech," in *Innovations in Information Technology, 2006*, November 2006, pp. 1–5.

[10] G. Wei, "Discrete singular convolution for beam analysis," *Engineering Structures*, vol. 23, pp. 1045–1053, January 2001.

[11] J. Fung, "Literature Survey EE 381K Multi-Dimensional Signal Processing - Effects of Steering Delay Quantization in Beamforming," 10000 Burnet Road, Austin, Texas 78758, Applied Research Laboratories, The University of Texas at Austin.

[12] J. Li and S. P., *Robust Adaptive Beamforming (1st edition)*. John Wiley and Sons, Inc, October 2005.

[13] B. Gillespie, H. Malvar, and D. Florencio, "Speech Dereverberation via Maximum-Kurtosis Subband Adaptive Filtering," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 6. Salt Lake City, UT , USA: IEEE, May 2001, pp. 3701–3704.

[14] Y. Liu and Q. Wan, "Digital Ultra Wideband Beamformer based on Minimum Variance Multi-Frequency Distortionless Restriction," May 2010.

[15] J. Gu and J. Wolfe, "Robust Adaptive Beamforming using Variable Loading," in *Sensor Array and Multichannel Processing, 2006. Fourth IEEE Workshop*, July 2006, pp. 1–5.

[16] J. Li, P. Stoica, and Z. Wang, "On Robust Capon Beamforming and Diagonal Loading," *Signal Processing, IEEE Transactions*, vol. 51, no. 7, pp. 1702–1715, July 2003.

[17] P. Lilja and H. Saarnisaari, "Robust Adaptive Beamforming in Software Defined Radio with Adaptive Diagonal Loading," in *Military Communications Conference, 2005. MILCOM 2005. IEEE*, vol. 4, October 2005, pp. 2596–2601.

[18] T. S. Laseetha and R. Sukanesh, "Robust Adaptive Beamformers using Diagonal Loading," *Cyber Journals: Multidisciplinary Journals in Science and Technology, Journal of Selected Areas in Telecommunications (JSAT), March Edition 2011*, pp. 73–79, March 2011.

[19] E. Mabande, A. Schad, and W. Kellermann, "Design of Robust Superdirective Beamformers as a Convex Optimization Problem," in *Acoustics, Speech and Signal Processing, 2009. ICASSP 2009. IEEE International Conference*, April 2009, pp. 77 –80.

[20] M. Wölfel and J. McDonough, *Distant Speech Recognition*. John Wiley and Sons, Inc, June 2009.

[21] M. Grant and S. Boyd, *cvx Users' Guide for cvx Version 1.21, Code Commit: 808, 2011-07-17, Doc Commit: 806, 2011-02-25*, d/b/a CVX Research, 1104 Claire Ave., Austin, TX 78703-2502, April 2011.

[22] P. Tsai, K. Ebrahim, G. Lange, Y. Paichard, and M. Inggs, "Null placement in a circular antenna array for passive coherent location systems," in *Radar Conference, 2010 IEEE*, May 2010, pp. 1140–1143.

[23] W. Herbordt and W. Kellermann, "Computationally efficient frequency-domain robust generalized sidelobe canceller," in *7th International Workshop on Acoustic Echo and Noise Control*, Darmstadt University of Technology, Darmstadt, Germany, 2001.

[24] J. Shynk, "Frequency-Domain and Multirate Adaptive Filtering," *Signal Processing Magazine, IEEE*, vol. 9, no. 1, pp. 14 –37, January 1992.

[25] J. Chen, H. Gu, H. Wang, and W. Su, "Mathematical analysis of main-to-sidelobe ratio after pulse compression in pseudorandom code phase modulation cw radar," in *Radar Conference 2008, IEEE*, May 2008, pp. 1–5.

[26] G. W. Elko, "A new technique to measure electroacoustic transducer directivity indices in reverberant fields," in *Applications of Signal Processing to Audio and Acoustics, 1993. Final Program and Paper Summaries, 1993 IEEE Workshop*, October 1993, pp. 64–67.

[27] H. Yi and P. Loizou, "Evaluation of objective quality measures for speech enhancement," *Audio, Speech, and Language Processing, IEEE Transactions*, vol. 16, no. 1, pp. 229–238, January 2008.

[28] J. H. L. Hansen and B. L. Pellom, "An effective quality evaluation protocol for speech enhancement algorithms," in *The International Conference on Speech and Language Processing*, 1998, pp. 2819–2822.

[29] D. Klatt, "Prediction of perceived phonetic distance from critical-band spectra: A first step," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '82.*, vol. 7, May 1982, pp. 1278–1281.

[30] A. W. Rix, J. G. Beerends, M. P. Hollier, and A. P. 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)*, vol. 2, 2001, pp. 749–752.

[31] S. R. Quackenbush, T. P. Barnwell, and M. A. Clements, *Objective Measures of Speech Quality (Prentice-Hall signal processing series) (1st edition)*. Prentice Hall, 1988.

[32] N. Kitawaki, H. Nagabuchi, and K. Itoh, "Objective quality evaluation for low-bit-rate speech coding systems," *Selected Areas in Communications, IEEE Journal*, vol. 6, no. 2, pp. 242–248, February 1988.

[33] J. Tribolet, P. Noll, B. McDermott, and R. Crochiere, "A study of complexity and quality of speech waveform coders," in *Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '78*, vol. 3, April 1978, pp. 586–590.

[34] P. C. Loizou, *Speech enhancement: Theory and Practice (Signal Processing and Communication), 1 edition*. CRC Press, June 2007.

[35] R. Crochiere, J. Tribolet, and L. Rabiner, "An interpretation of the log-likelihood ratio as a measure of waveform coder performance," *Acoustics, Speech and Signal Processing, IEEE Transactions*, vol. 28, no. 3, pp. 318–323, June 1980.

[36] B. Grundlehner, J. Lecocq, R. Balan, and J. Rosca, "Performance assessment method for speech enhancement systems," in *SPS-DARTS 2005, The first annual IEEE BENELUX/DSP Valley Signal Processing Symposium*, Het Provinciehuis, Antwerp, Belgium, April 2005.

[37] T. Rohdenburg, S. Goetze, V. Hohmann, K.-D. Kammeyer, and B. Kollmeier, "Objective perceptual quality assessment for self-steering binaural hearing aid microphone arrays," in *IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP), Las Vegas, USA*, March-April 2008, pp. 2449–2452.

[38] A. E. Conway, "Output-based method of applying pesq to measure the perceptual quality of framed speech signals," in *Wireless Communications and Networking Conference, 2004. WCNC. 2004 IEEE*, vol. 4, March 2004, pp. 2521–2526.

[39] A. Technologies, *PESQ Test Application: An objective audio test solution with the E5515C*, http://cp.literature.agilent.com/litweb/pdf/5990-5961EN.pdf, 2010.

[40] S. Young, G. Evermann, M. Gales, T. Hain, D. Kershaw, X. Liu, G. Moore, J. Odell, D. Ollason, D. Povey, V. Valtchev, and P. Woodland, *The HTK Book (V. 3.4)*. Cambridge University Engineering Department, December 2006.

[41] K. Vertanen, "Baseline WSJ acoustic models for HTK and Sphinx: Training recipes and recognition experiments," Cavendish Laboratory, University of Cambridge, Tech. Rep., 2006.

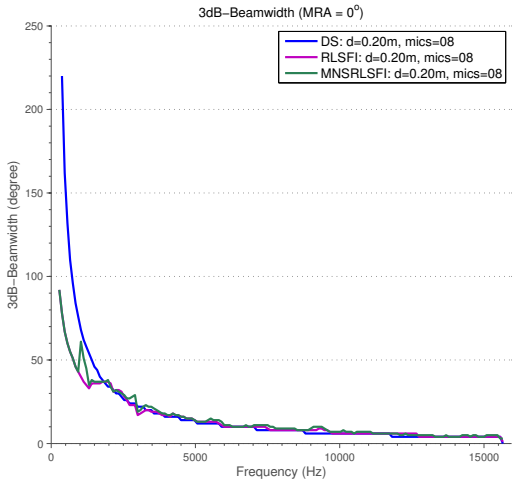
[42] S. Haykin and K. J. R. Liu, *Handbook on Array Processing and Sensor Networks*. John Wiley and Sons, Inc., January 2009.

[43] M. Bar, "Visual objects in context," *Nature Reviews Neuroscience*, vol. 5, pp. 617–629, August 2004.

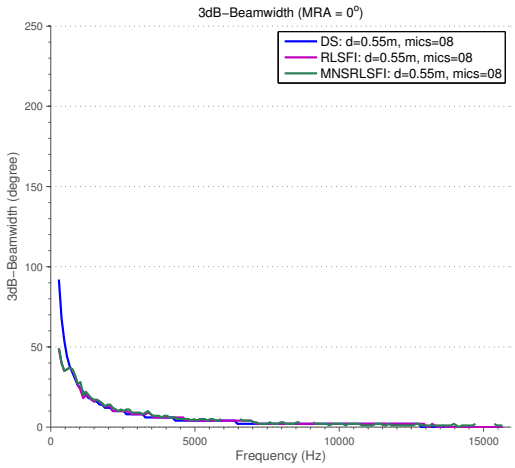
[44] R. Scholte, B. Roozen, and I. Lopez, "Twelfth international congress on sound and vibration," Portugal, Lisbon, July 2005.

[45] T. Habib and H. Romsdorfer, "Comparison of SRP-PHAT and Multiband-PoPi Algorithms for Speaker Localization using Particle Filters," in *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, September 2010.

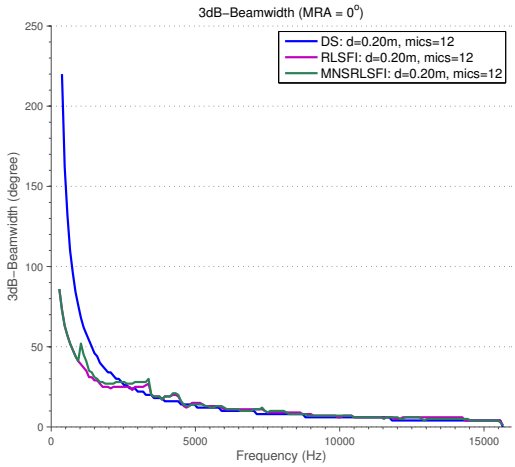
A
Graphical Results (Figure)



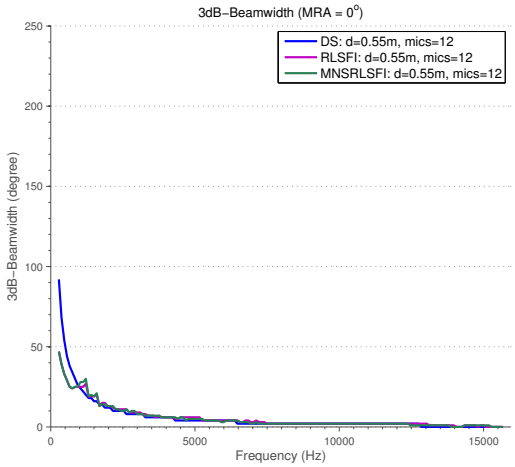
(a) 3dB-BW over all frequencies (d=0.20m, #Mics=08).



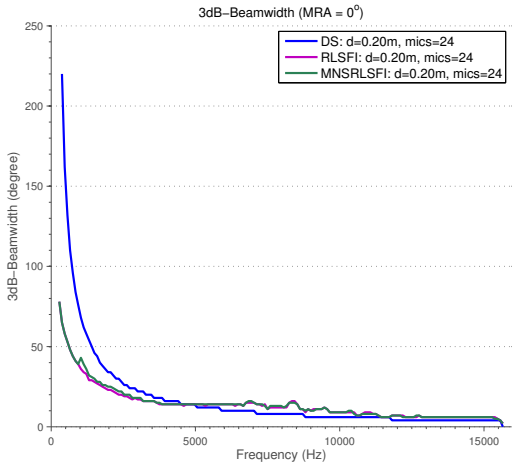
(b) 3dB-BW over all frequencies (d=0.55m, #Mics=08).



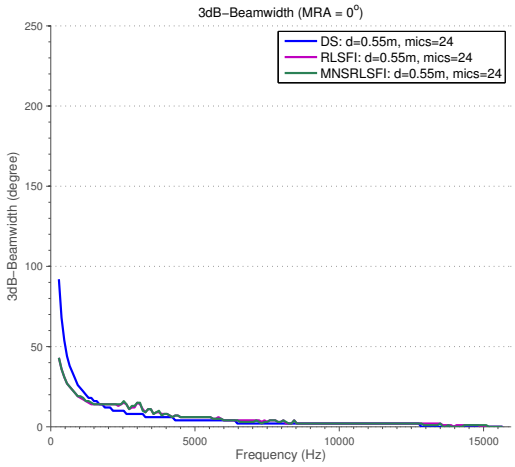
(c) 3dB-BW over all frequencies (d=0.20m, #Mics=12).



(d) 3dB-BW over all frequencies (d=0.55m, #Mics=12).

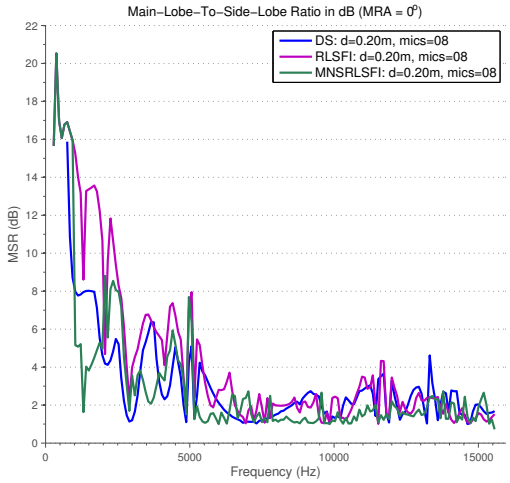


(e) 3dB-BW over all frequencies (d=0.20m, #Mics=24).

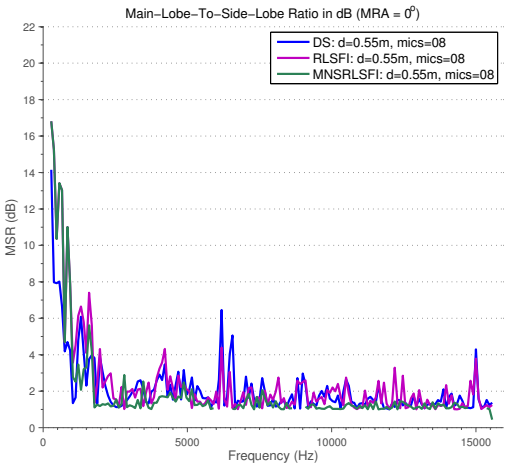


(f) 3dB-BW over all frequencies (d=0.55m, #Mics=24).

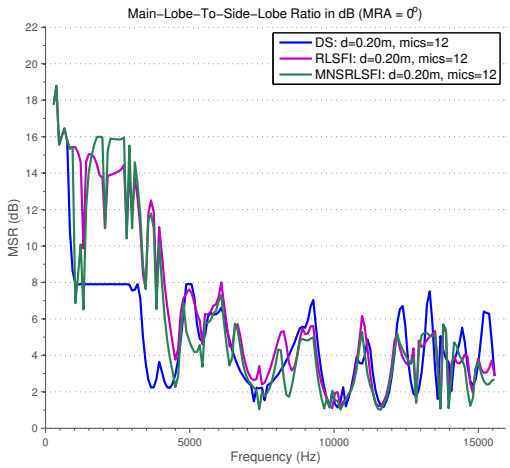
Figure A.1: These plots depict the 3dB-BW.



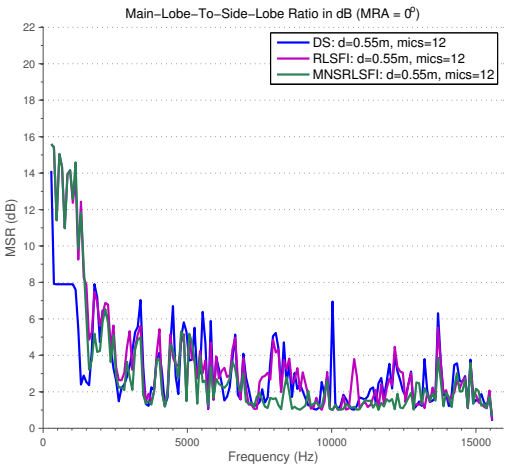
(a) MSR over all frequencies(d=0.20m, #Mics=08).



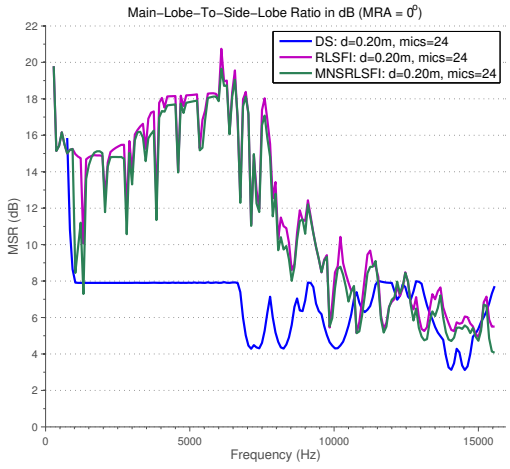
(b) MSR over all frequencies(d=0.55m, #Mics=08).



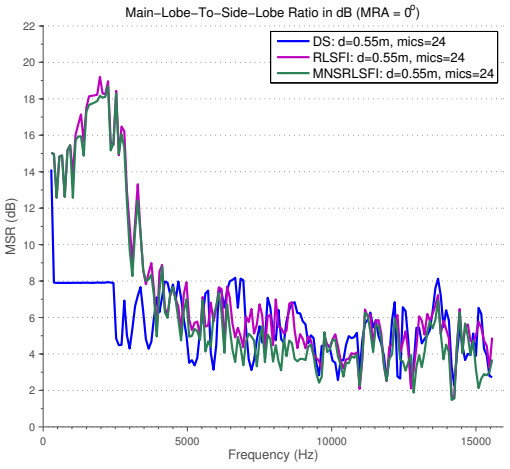
(c) MSR over all frequencies(d=0.20m, #Mics=12).



(d) MSR over all frequencies(d=0.55m, #Mics=12).

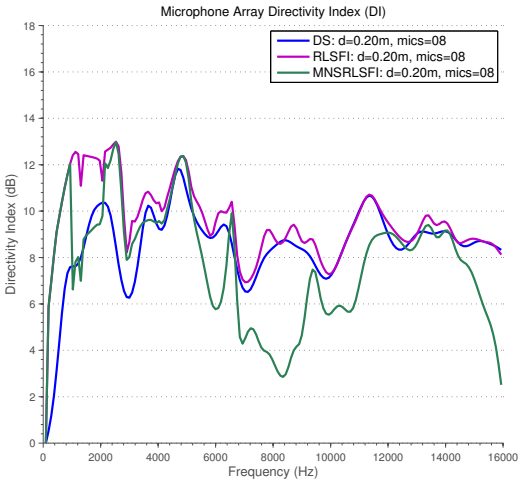


(e) MSR over all frequencies(d=0.20m, #Mics=24).

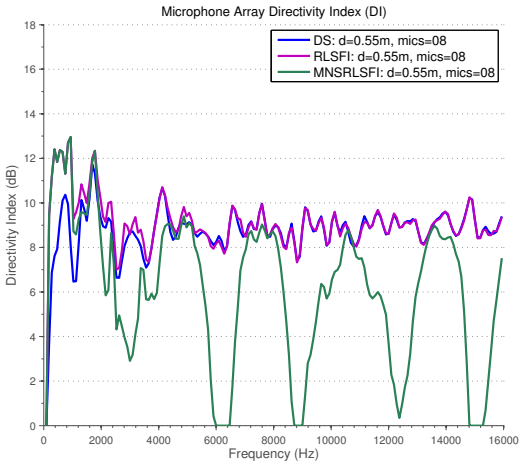


(f) MSR over all frequencies(d=0.55m, #Mics=24).

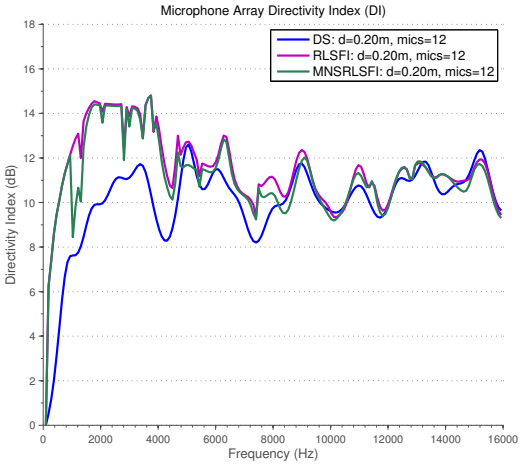
Figure A.2: These plots depict the MSR.



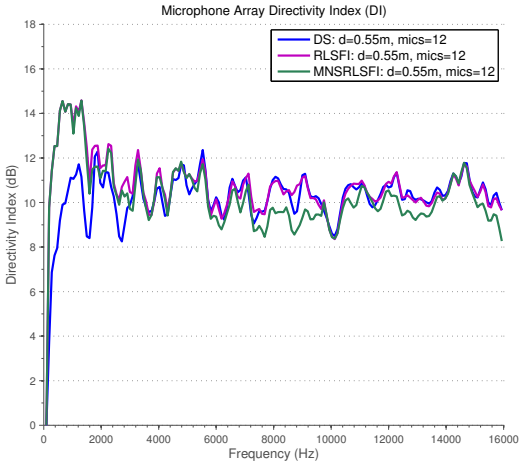
(a) DI over all frequencies(d=0.20m, #Mics=08).



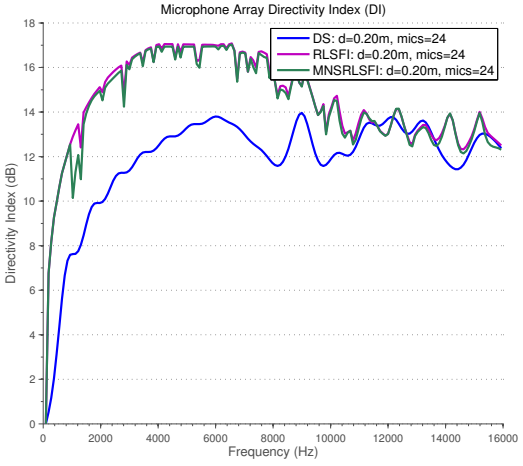
(b) DI over all frequencies(d=0.55m, #Mics=08).



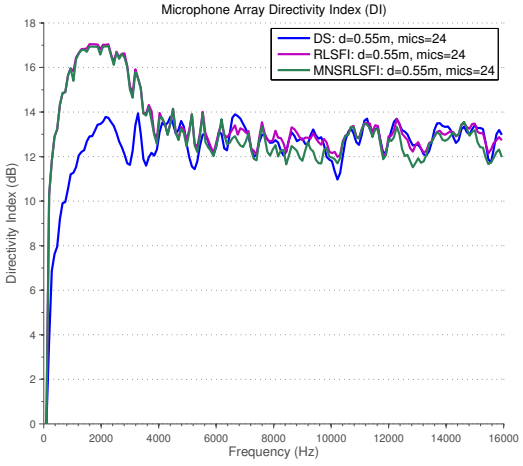
(c) DI over all frequencies(d=0.20m, #Mics=12).



(d) DI over all frequencies(d=0.55m, #Mics=12).

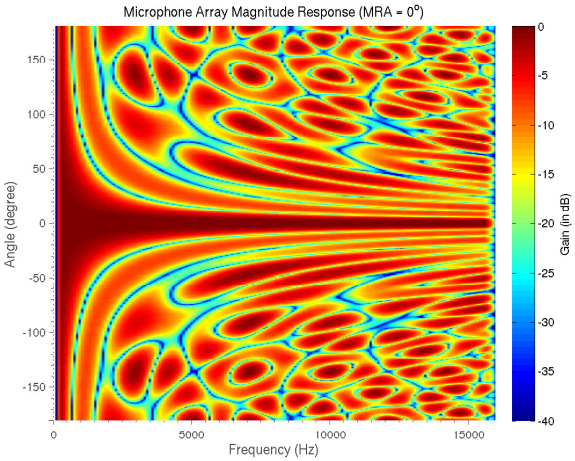


(e) DI over all frequencies(d=0.20m, #Mics=24).

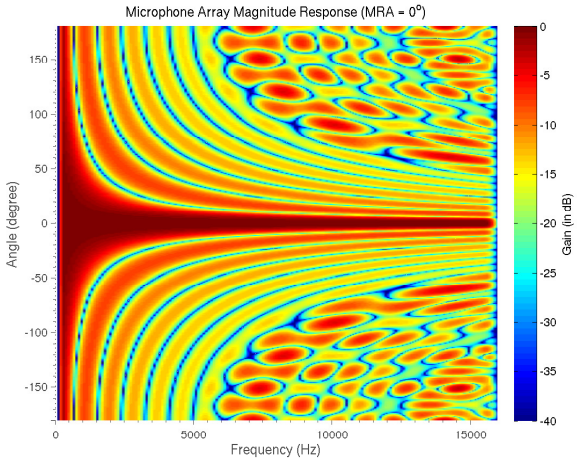


(f) DI over all frequencies(d=0.55m, #Mics=24).

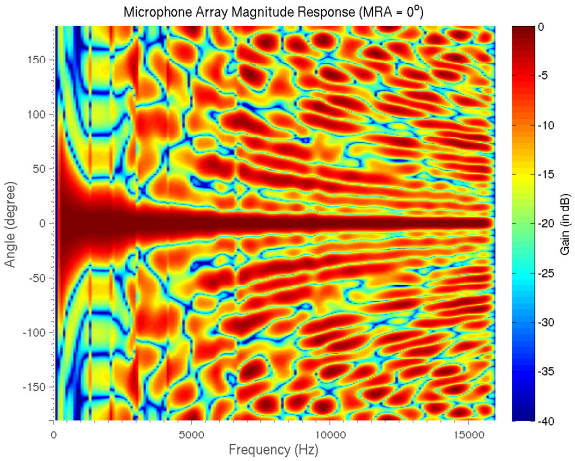
Figure A.3: These plots depict the DI.



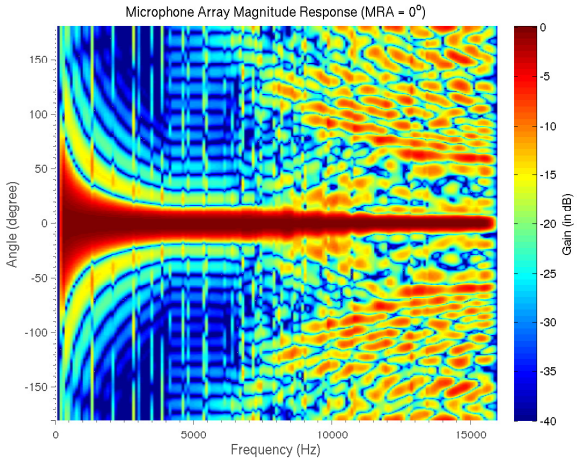
(a) Beampattern of the Delay&Sum-BF. (d=0.20m, #Mics=08).



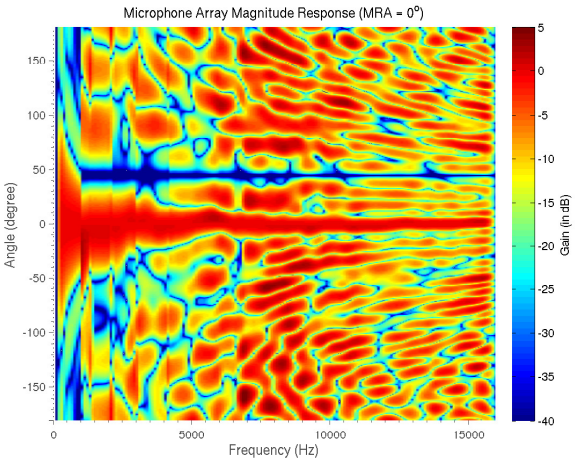
(b) Beampattern of the Delay&Sum-BF (d=0.20m, #Mics=24).



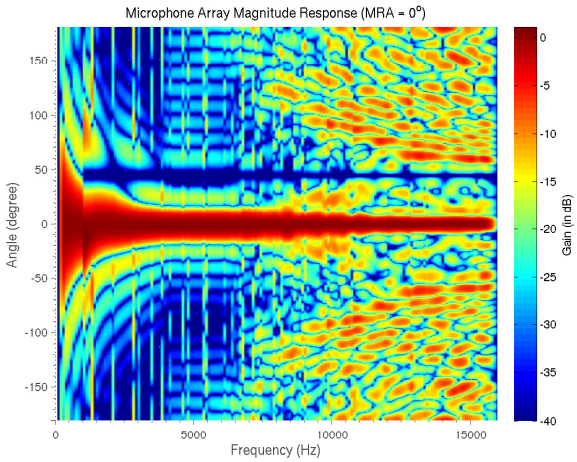
(c) Beampattern of the RLSFI-BF (d=0.20m, #Mics=08).



(d) Beampattern of the RLSFI-BF (d=0.20m, #Mics=24).

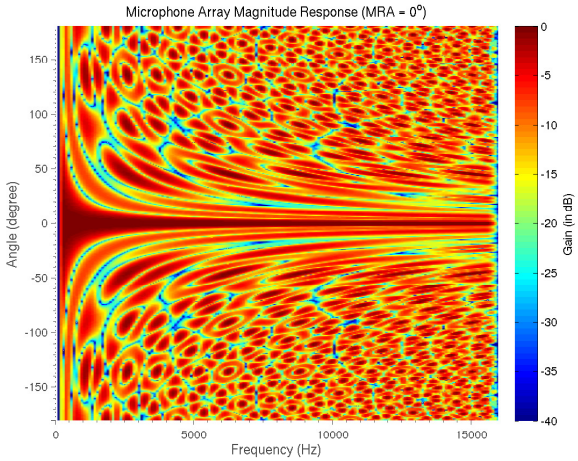


(e) Beampattern of the MNSRLSFI-BF (d=0.20m, #Mics=08).

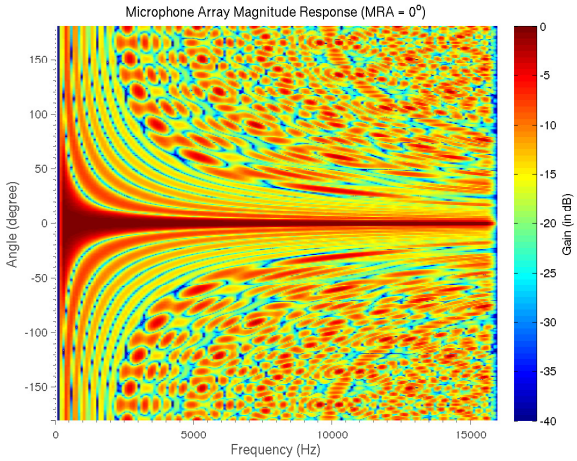


(f) Beampattern of the MNSRLSFI-BF (d=0.20m, #Mics=24).

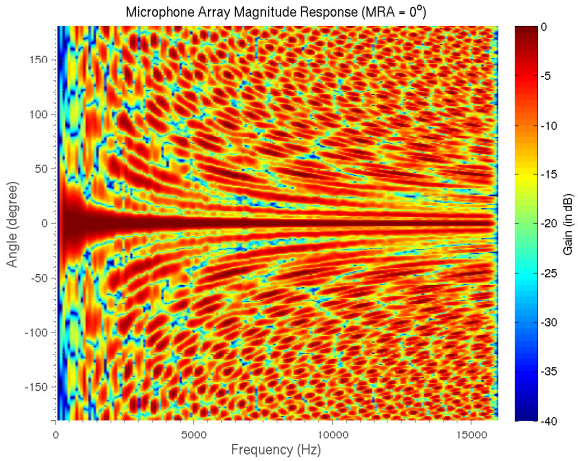
Figure A.4: These plots depict beampatterns of different beamformers for d=0.20m.



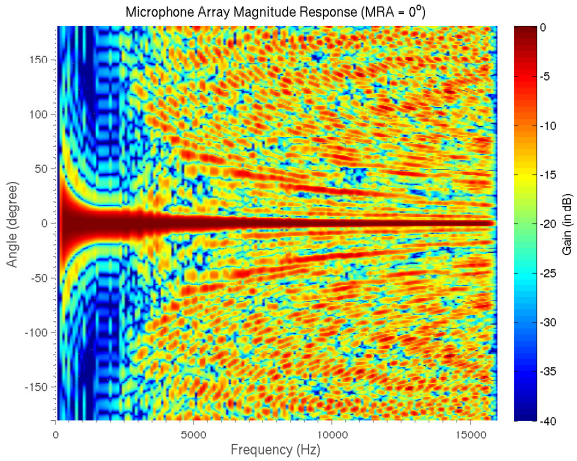
(a) Beampattern of the Delay&Sum-BF. (d=0.55m, #Mics=08).



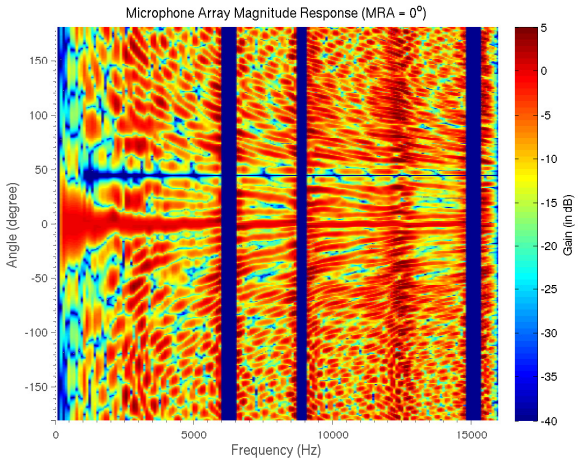
(b) Beampattern of the Delay&Sum-BF (d=0.55m, #Mics=24).



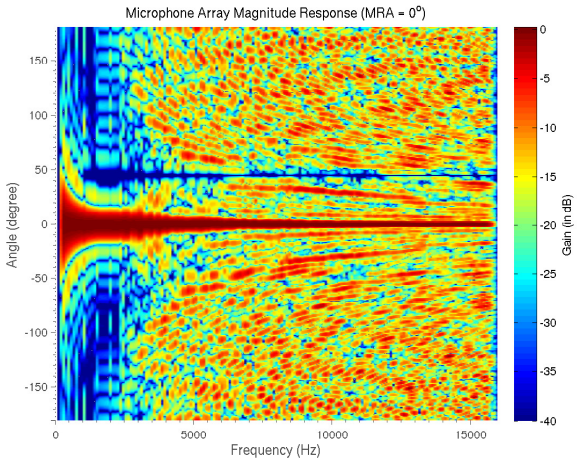
(c) Beampattern of the RLSFI-BF (d=0.55m, #Mics=08).



(d) Beampattern of the RLSFI-BF (d=0.55m, #Mics=24).

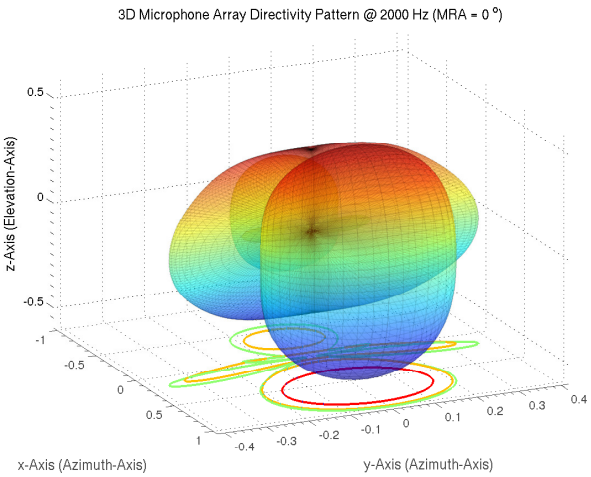


(e) Beampattern of the MNSRLSFI-BF (d=0.55m, #Mics=08).

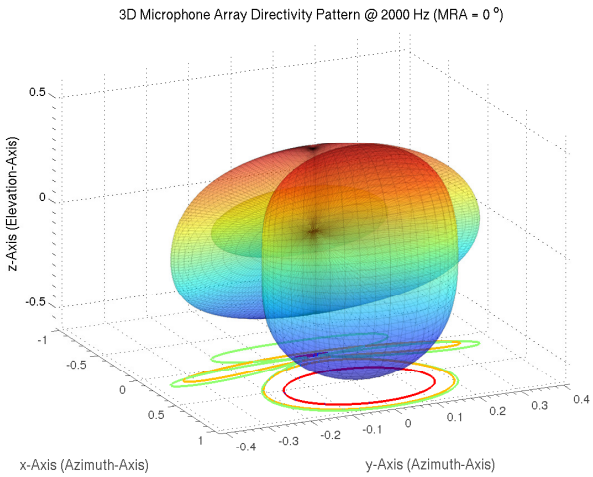


(f) Beampattern of the MNSRLSFI-BF (d=0.55m, #Mics=24).

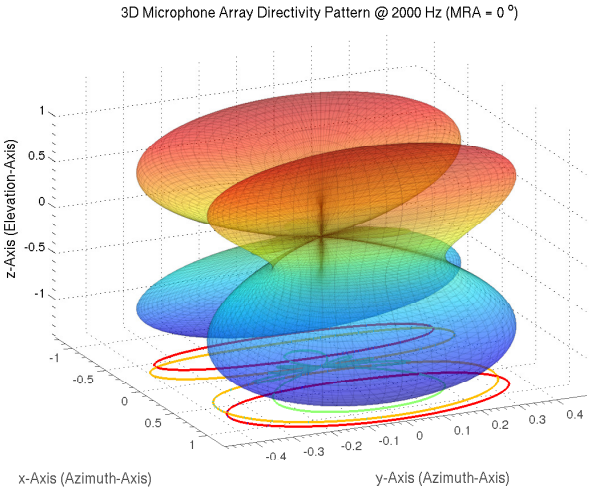
Figure A.5: These plots depict beampatterns of different beamformers for d=0.50m.



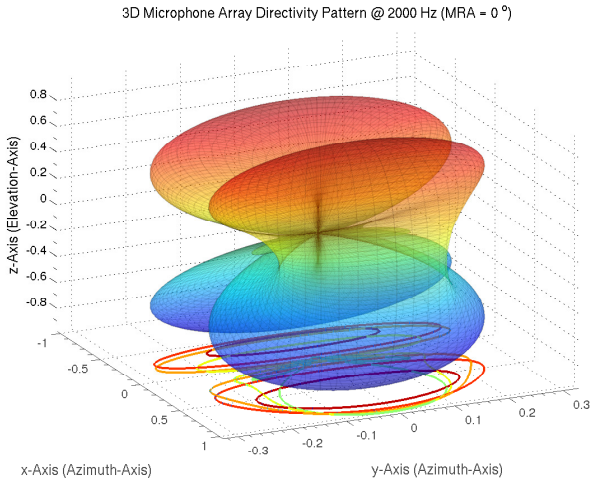
(a) 3D-Beampattern of the Delay&Sum-BF. (d=0.20m, #Mics=08).



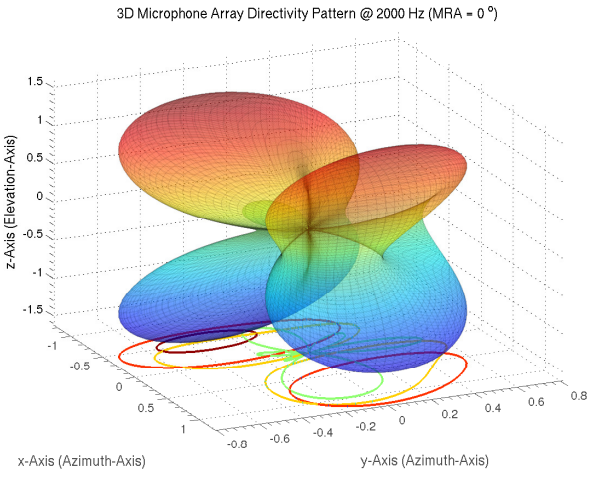
(b) 3D-Beampattern of the Delay&Sum-BF (d=0.20m, #Mics=24).



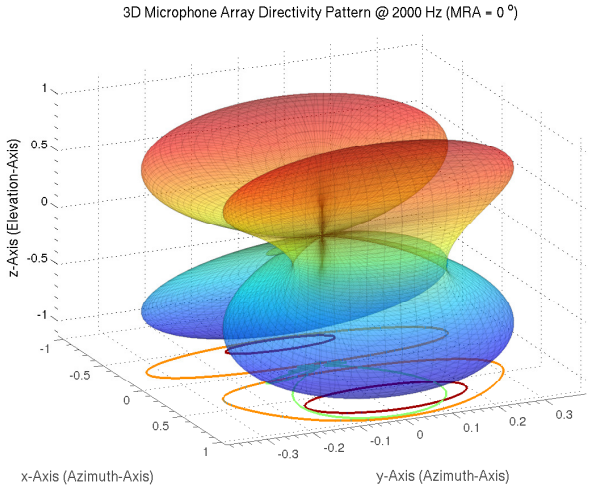
(c) 3D-Beampattern of the RLSFI-BF (d=0.20m, #Mics=08).



(d) 3D-Beampattern of the RLSFI-BF (d=0.20m, #Mics=24).

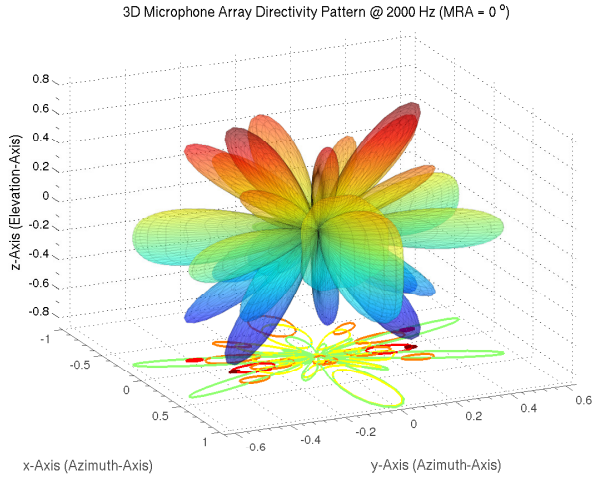


(e) 3D-Beampattern of the MNSRLSFI-BF (d=0.20m, #Mics=08).

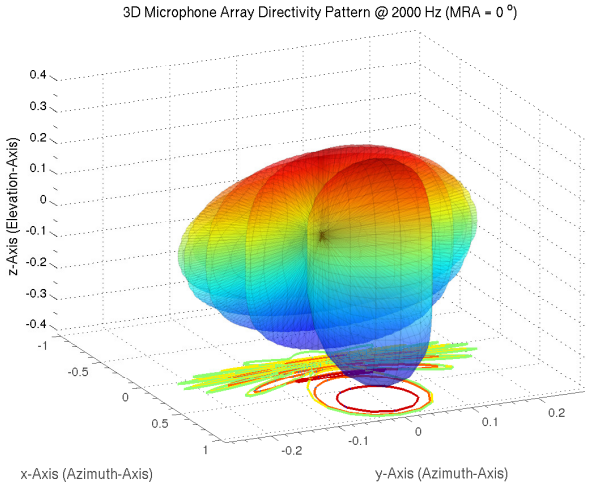


(f) 3D-Beampattern of the MNSRLSFI-BF (d=0.20m, #Mics=24).

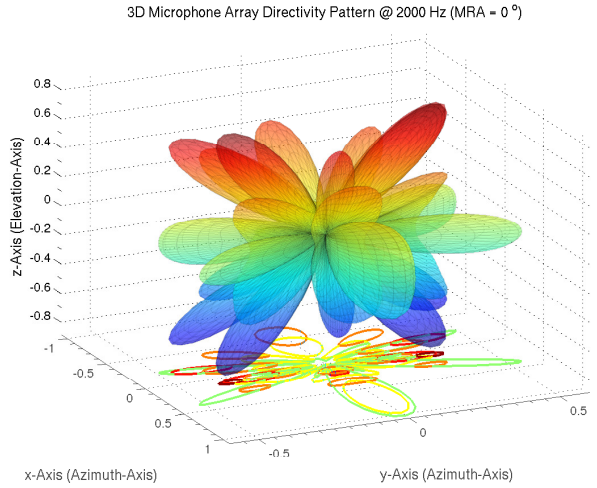
Figure A.6: These plots depict beampatterns of different beamformers for d=0.20m and f=2000Hz.



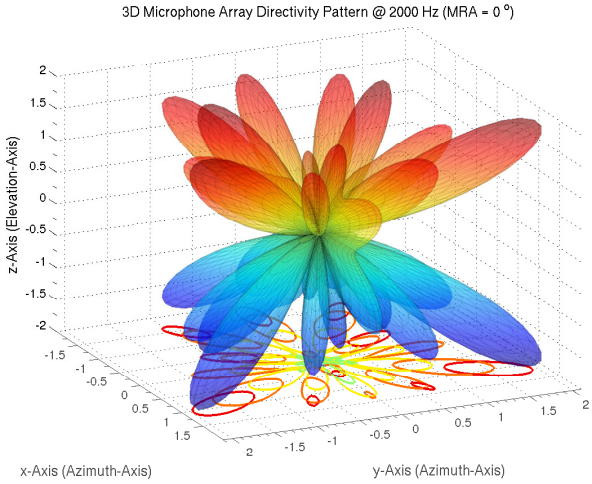
(a) 3D-Beampattern of the Delay&Sum-BF. (d=0.55m, #Mics=08).



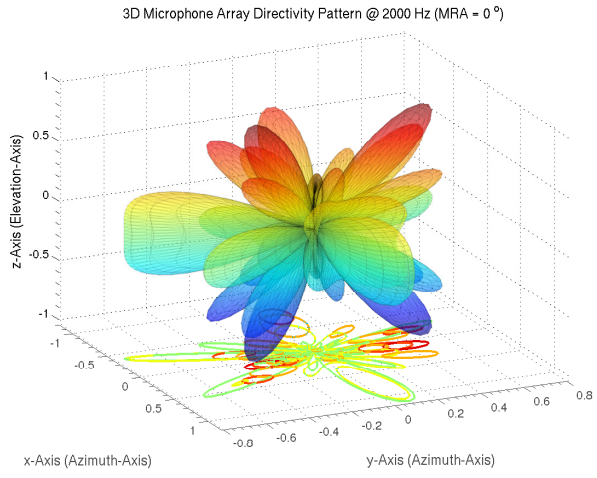
(b) 3D-Beampattern of the Delay&Sum-BF (d=0.55m, #Mics=24).



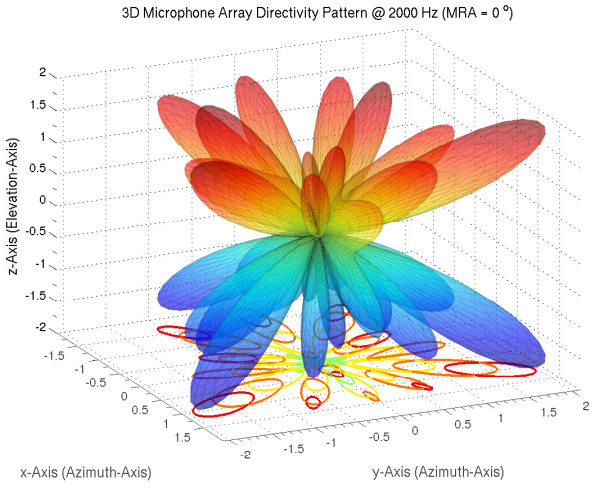
(c) 3D-Beampattern of the RLSFI-BF (d=0.55m, #Mics=08).



(d) 3D-Beampattern of the RLSFI-BF (d=0.55m, #Mics=24).



(e) 3D-Beampattern of the MNSRLSFI-BF (d=0.55m, #Mics=08).



(f) 3D-Beampattern of the MNSRLSFI-BF (d=0.55m, #Mics=24).

Figure A.7: These plots depict beampatterns of different beamformers for d=0.50m and f=2000Hz.

B
Numerical Results (Tables)

B.0.1 Results based on Synthetic Data

Tables of 1^{st} Scenario (Speaker 03 {male} and 12 {male} @ MRA=$\{00^\circ, 45^\circ\}$)

DS-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**74.17 %**	50.83 %	12.467 dB	-1.234 dB	2.414	-4.745	51.640	2.340	1.555	2.092	1.880
(12,20)	**71.67 %**	50.83 %	12.563 dB	-1.251 dB	2.540	-4.719	51.631	2.342	1.427	2.095	1.817
(24,20)	**71.67 %**	50.83 %	12.578 dB	-1.242 dB	2.555	-4.721	51.588	2.346	1.414	2.097	1.813
(08,55)	**77.50 %**	50.00 %	11.950 dB	-3.407 dB	2.565	-4.359	53.168	2.359	1.397	2.115	1.807
(12,55)	**81.67 %**	50.00 %	12.068 dB	-3.427 dB	2.606	-4.324	52.801	2.386	1.374	2.132	1.810
(24,55)	**84.17 %**	50.00 %	12.099 dB	-3.427 dB	2.627	-4.319	52.590	2.388	1.357	2.135	1.803

Table B.1: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

DS-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**58.33 %**	46.67 %	1.732 dB	-1.074 dB	2.382	-4.626	50.527	2.292	1.569	2.085	1.866
(12,20)	**60.83 %**	46.67 %	1.726 dB	-1.072 dB	2.514	-4.834	50.511	2.241	1.403	2.047	1.757
(24,20)	**60.83 %**	46.67 %	1.730 dB	-1.079 dB	2.547	-4.606	50.408	2.284	1.396	2.083	1.776
(08,55)	**70.00 %**	45.83 %	2.586 dB	-3.045 dB	2.587	-4.311	52.068	2.343	1.375	2.118	1.791
(12,55)	**78.33 %**	48.33 %	2.577 dB	-3.070 dB	2.552	-4.530	51.566	2.284	1.380	2.079	1.765
(24,55)	**77.50 %**	45.83 %	2.595 dB	-3.072 dB	2.648	-4.264	51.298	2.378	1.341	2.143	1.794

Table B.2: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45^\circ$.

RLSFI-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.83 %**	50.83 %	14.255 dB	-3.887 dB	2.543	-4.106	52.561	2.374	1.435	2.142	1.835
(12,20)	**83.33 %**	50.83 %	14.463 dB	-4.212 dB	2.570	-4.007	51.527	2.381	1.420	2.159	1.834
(24,20)	**85.00 %**	50.83 %	13.923 dB	-4.471 dB	2.475	-3.932	51.513	2.388	1.522	2.167	1.888
(08,55)	**80.00 %**	50.00 %	15.793 dB	-2.768 dB	2.579	-3.947	58.129	2.432	1.382	2.141	1.824
(12,55)	**79.17 %**	50.00 %	15.792 dB	-3.224 dB	2.633	-3.837	57.158	2.414	1.325	2.146	1.789
(24,55)	**80.83 %**	50.00 %	16.231 dB	-2.582 dB	2.642	-3.911	60.362	2.489	1.332	2.155	1.823

Table B.3: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

RLSFI-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**70.83 %**	46.67 %	2.201 dB	-3.030 dB	2.455	-4.048	50.590	2.320	1.510	2.134	1.850
(12,20)	**71.67 %**	46.67 %	2.278 dB	-3.139 dB	2.525	-4.277	50.452	2.297	1.425	2.109	1.797
(24,20)	**77.50 %**	46.67 %	2.370 dB	-3.491 dB	2.353	-3.910	50.701	2.349	1.631	2.156	1.925
(08,55)	**74.17 %**	45.83 %	2.755 dB	-3.018 dB	2.579	-3.956	54.868	2.392	1.388	2.144	1.815
(12,55)	**73.33 %**	48.33 %	2.754 dB	-3.344 dB	2.571	-4.179	55.696	2.333	1.353	2.096	1.766
(24,55)	**71.67 %**	45.83 %	2.566 dB	-2.379 dB	2.631	-4.019	56.387	2.419	1.337	2.143	1.800

Table B.4: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45^\circ$.

MNS-RLSFI-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**83.33 %**	50.83 %	14.474 dB	-3.899 dB	2.493	-4.062	51.267	2.372	1.497	2.153	1.868
(12,20)	**84.17 %**	50.83 %	14.490 dB	-4.210 dB	2.565	-3.989	51.049	2.378	1.428	2.162	1.838
(24,20)	**82.50 %**	50.83 %	13.909 dB	-4.465 dB	2.460	-3.935	51.502	2.394	1.541	2.170	1.901
(08,55)	**81.67 %**	50.00 %	16.132 dB	-2.790 dB	2.637	-3.883	57.925	2.463	1.344	2.161	1.821
(12,55)	**86.67 %**	50.00 %	15.880 dB	-3.243 dB	2.618	-3.799	58.652	2.437	1.341	2.149	1.805
(24,55)	**83.33 %**	50.00 %	16.327 dB	-2.581 dB	2.616	-3.892	59.952	2.465	1.348	2.147	1.819

Table B.5: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

MNS-RLSFI-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**79.17 %**	46.67 %	2.194 dB	-3.038 dB	2.456	-3.998	50.579	2.328	1.515	2.141	1.857
(12,20)	**79.17 %**	46.67 %	2.280 dB	-3.135 dB	2.514	-4.244	50.800	2.277	1.422	2.099	1.784
(24,20)	**79.17 %**	46.67 %	2.371 dB	-3.477 dB	2.344	-3.922	50.846	2.323	1.625	2.142	1.909
(08,55)	**81.67 %**	45.83 %	2.758 dB	-3.053 dB	2.666	-3.877	54.620	2.421	1.318	2.165	1.796
(12,55)	**80.00 %**	48.33 %	2.763 dB	-3.373 dB	2.584	-4.106	55.188	2.333	1.344	2.104	1.763
(24,55)	**75.83 %**	45.83 %	2.566 dB	-2.375 dB	2.663	-3.999	56.421	2.411	1.298	2.139	1.776

Table B.6: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45^\circ$.

MPDRDL-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**76.67 %**	50.83 %	19.601 dB	9.088 dB	8.801	-7.967	29.989	3.222	-4.290	2.462	-0.528
(12,20)	**75.00 %**	50.83 %	20.223 dB	8.836 dB	4.232	-7.804	26.709	3.242	0.452	2.505	1.850
(24,20)	**71.67 %**	50.83 %	22.641 dB	9.098 dB	0.909	-7.762	25.433	3.241	3.883	2.516	3.559
(08,55)	**66.67 %**	50.00 %	15.718 dB	0.184 dB	3.641	-6.551	32.322	2.676	0.669	2.274	1.658
(12,55)	**63.33 %**	50.00 %	16.374 dB	-0.479 dB	0.765	-6.253	30.250	2.678	3.648	2.308	3.146
(24,55)	**64.17 %**	50.00 %	18.302 dB	-0.965 dB	0.632	-6.077	32.511	2.614	3.727	2.273	3.147

Table B.7: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0°$.

MPDRDL-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.83 %**	46.67 %	3.356 dB	9.590 dB	8.966	-7.780	32.446	3.293	-4.440	2.491	-0.573
(12,20)	**77.50 %**	46.67 %	3.320 dB	10.228 dB	5.671	-7.658	28.561	3.337	-0.988	2.547	1.177
(24,20)	**72.50 %**	46.67 %	3.400 dB	9.578 dB	1.152	-7.433	27.872	3.321	3.660	2.558	3.483
(08,55)	**59.17 %**	45.83 %	3.119 dB	0.930 dB	3.989	-6.605	33.397	2.928	0.453	2.383	1.674
(12,55)	**62.50 %**	48.33 %	3.733 dB	-0.844 dB	0.973	-6.204	33.368	2.794	3.476	2.345	3.111
(24,55)	**64.17 %**	45.83 %	3.604 dB	-0.876 dB	0.615	-6.108	34.379	2.832	3.858	2.362	3.318

Table B.8: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

MPDRVL-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	0.00 %	50.83 %	10.036 dB	-1.490 dB	0.813	-4.585	32.875	2.073	3.210	2.106	2.616
(12,20)	35.00 %	50.83 %	10.734 dB	-1.569 dB	0.751	-4.388	33.659	2.157	3.318	2.153	2.710
(24,20)	34.17 %	50.83 %	12.016 dB	-1.779 dB	0.733	-4.285	32.243	2.160	3.351	2.171	2.732
(08,55)	0.00 %	50.00 %	8.383 dB	-1.955 dB	0.693	-5.237	37.553	2.053	3.280	2.023	2.629
(12,55)	38.33 %	50.00 %	7.832 dB	-3.148 dB	0.689	-4.992	37.410	2.023	3.267	2.025	2.608
(24,55)	33.33 %	50.00 %	8.260 dB	-3.474 dB	0.691	-4.921	40.802	2.129	3.298	2.056	2.668

Table B.9: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0°$.

MPDRVL-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	31.67 %	46.67 %	1.910 dB	-0.795 dB	0.672	-5.315	34.003	2.239	3.445	2.131	2.814
(12,20)	32.50 %	46.67 %	2.134 dB	-0.692 dB	0.619	-5.237	35.264	2.253	3.497	2.134	2.844
(24,20)	31.67 %	46.67 %	1.917 dB	-0.985 dB	0.748	-5.101	35.215	2.309	3.399	2.170	2.824
(08,55)	34.17 %	45.83 %	1.456 dB	-1.454 dB	0.580	-5.633	42.374	2.180	3.429	2.024	2.755
(12,55)	36.67 %	48.33 %	1.743 dB	-2.570 dB	0.586	-5.532	38.586	2.128	3.425	2.032	2.736
(24,55)	33.33 %	45.83 %	1.767 dB	-2.987 dB	0.621	-5.369	42.449	2.218	3.409	2.059	2.764

Table B.10: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

GSC (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**71.67 %**	50.83 %	12.467 dB	-1.234 dB	0.768	-4.994	21.577	2.784	3.788	2.499	3.291
(12,20)	**73.33 %**	50.83 %	12.563 dB	-1.251 dB	0.716	-5.002	21.994	2.739	3.810	2.474	3.278
(24,20)	**69.17 %**	50.83 %	12.578 dB	-1.242 dB	0.989	-5.243	22.942	2.649	3.466	2.409	3.059
(08,55)	**77.50 %**	50.00 %	11.950 dB	-3.407 dB	0.775	-4.833	21.638	2.954	3.883	2.590	3.424
(12,55)	**82.50 %**	50.00 %	12.068 dB	-3.427 dB	0.686	-4.864	21.287	2.977	3.990	2.601	3.490
(24,55)	**83.33 %**	50.00 %	12.099 dB	-3.427 dB	0.915	-4.933	22.447	2.923	3.712	2.563	3.322

Table B.11: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0°$.

GSC (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**61.67 %**	46.67 %	1.732 dB	-1.074 dB	0.753	-5.359	23.686	2.792	3.789	2.465	3.291
(12,20)	**65.83 %**	46.67 %	1.726 dB	-1.072 dB	0.707	-5.438	24.040	2.779	3.825	2.452	3.301
(24,20)	**64.17 %**	46.67 %	1.730 dB	-1.079 dB	0.879	-5.492	24.706	2.715	3.603	2.413	3.156
(08,55)	**67.50 %**	45.83 %	2.586 dB	-3.045 dB	0.767	-5.162	26.001	2.974	3.862	2.548	3.413
(12,55)	**79.17 %**	48.33 %	2.577 dB	-3.070 dB	0.664	-5.109	24.031	3.039	4.026	2.597	3.532
(24,55)	**77.50 %**	45.83 %	2.595 dB	-3.072 dB	1.021	-5.398	25.686	3.019	3.631	2.557	3.322

Table B.12: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

Tables of 2^{nd} Scenario (Speaker 03 {male} and 12 {male} @ MRA=$\{300°, 120°\}$)

DS-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**84.17 %**	50.83 %	11.148 dB	-2.842 dB	0.579	-3.968	15.627	3.121	4.238	2.766	3.700
(12,20)	**80.83 %**	50.83 %	11.107 dB	-2.829 dB	0.718	-3.989	15.716	3.099	4.082	2.754	3.611
(24,20)	**82.50 %**	50.83 %	11.168 dB	-2.844 dB	0.328	-3.949	15.139	3.131	4.507	2.776	3.840
(08,55)	**84.17 %**	51.67 %	11.866 dB	-3.812 dB	0.627	-5.093	16.010	3.097	4.172	2.681	3.654
(12,55)	**77.50 %**	52.50 %	11.813 dB	-3.778 dB	0.574	-5.160	15.929	3.095	4.225	2.677	3.680
(24,55)	**85.00 %**	52.50 %	11.946 dB	-3.826 dB	0.376	-5.099	14.897	3.273	4.546	2.773	3.932

Table B.13: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

DS-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**75.00 %**	48.33 %	2.474 dB	-2.474 dB	0.694	-4.942	19.010	3.043	4.043	2.644	3.555
(12,20)	**79.17 %**	48.33 %	2.462 dB	-2.455 dB	0.745	-4.963	18.880	3.042	3.991	2.643	3.529
(24,20)	**82.50 %**	48.33 %	2.474 dB	-2.477 dB	0.429	-4.918	18.148	3.059	4.333	2.660	3.710
(08,55)	**80.00 %**	45.83 %	3.062 dB	-3.635 dB	0.623	-5.420	18.031	3.157	4.193	2.675	3.690
(12,55)	**75.00 %**	45.83 %	3.053 dB	-3.593 dB	0.651	-5.480	19.333	3.143	4.145	2.656	3.655
(24,55)	**87.50 %**	45.83 %	3.063 dB	-3.632 dB	0.346	-5.435	17.155	3.233	4.532	2.717	3.899

Table B.14: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

RLSFI-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**84.17 %**	50.83 %	11.148 dB	-2.842 dB	0.579	-3.968	15.627	3.121	4.238	2.766	3.700
(12,20)	**80.83 %**	50.83 %	11.107 dB	-2.829 dB	0.718	-3.989	15.716	3.099	4.082	2.754	3.611
(24,20)	**82.50 %**	50.83 %	11.168 dB	-2.844 dB	0.328	-3.949	15.139	3.131	4.507	2.776	3.840
(08,55)	**84.17 %**	51.67 %	11.866 dB	-3.812 dB	0.627	-5.093	16.010	3.097	4.172	2.681	3.654
(12,55)	**77.50 %**	52.50 %	11.813 dB	-3.778 dB	0.574	-5.160	15.929	3.095	4.225	2.677	3.680
(24,55)	**85.00 %**	52.50 %	11.946 dB	-3.826 dB	0.376	-5.099	14.897	3.273	4.546	2.773	3.932

Table B.15: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

RLSFI-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**75.00 %**	48.33 %	2.474 dB	-2.474 dB	0.694	-4.942	19.010	3.043	4.043	2.644	3.555
(12,20)	**79.17 %**	48.33 %	2.462 dB	-2.455 dB	0.745	-4.963	18.880	3.042	3.991	2.643	3.529
(24,20)	**82.50 %**	48.33 %	2.474 dB	-2.477 dB	0.429	-4.918	18.148	3.059	4.333	2.660	3.710
(08,55)	**80.00 %**	45.83 %	3.062 dB	-3.635 dB	0.623	-5.420	18.031	3.157	4.193	2.675	3.690
(12,55)	**75.00 %**	45.83 %	3.053 dB	-3.593 dB	0.651	-5.480	19.333	3.143	4.145	2.656	3.655
(24,55)	**87.50 %**	45.83 %	3.063 dB	-3.632 dB	0.346	-5.435	17.155	3.233	4.532	2.717	3.899

Table B.16: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MNS-RLSFI-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**85.83 %**	50.83 %	11.201 dB	-2.859 dB	0.264	-3.918	12.802	3.276	4.682	2.863	4.006
(12,20)	**86.67 %**	50.83 %	11.178 dB	-2.852 dB	0.174	-3.907	13.058	3.264	4.765	2.857	4.041
(24,20)	**85.83 %**	50.83 %	11.205 dB	-2.859 dB	0.162	-3.902	12.601	3.276	4.789	2.866	4.060
(08,55)	**89.17 %**	51.67 %	11.999 dB	-3.857 dB	0.254	-4.999	12.388	3.361	4.746	2.839	4.082
(12,55)	**87.50 %**	52.50 %	11.945 dB	-3.832 dB	0.258	-5.069	12.217	3.353	4.740	2.832	4.076
(24,55)	**89.17 %**	52.50 %	12.015 dB	-3.842 dB	0.206	-5.048	11.336	3.411	4.836	2.867	4.155

Table B.17: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

MNS-RLSFI-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**91.67 %**	48.33 %	2.480 dB	-2.481 dB	0.305	-4.903	16.143	3.117	4.513	2.702	3.834
(12,20)	**91.67 %**	48.33 %	2.461 dB	-2.478 dB	0.230	-4.893	16.183	3.112	4.588	2.700	3.869
(24,20)	**90.83 %**	48.33 %	2.475 dB	-2.485 dB	0.202	-4.889	15.940	3.117	4.622	2.704	3.888
(08,55)	**89.17 %**	45.83 %	3.081 dB	-3.664 dB	0.282	-5.342	15.719	3.278	4.638	2.754	3.978
(12,55)	**90.83 %**	45.83 %	3.056 dB	-3.635 dB	0.259	-5.404	15.388	3.263	4.655	2.745	3.980
(24,55)	**87.50 %**	45.83 %	3.068 dB	-3.644 dB	0.221	-5.388	14.837	3.282	4.711	2.760	4.019

Table B.18: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MPDRDL-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	0.00 %	50.83 %	19.348 dB	7.302 dB	0.993	-7.538	34.192	2.483	3.261	2.107	2.845
(12,20)	0.00 %	50.83 %	17.904 dB	6.954 dB	0.941	-7.497	34.564	2.477	3.307	2.104	2.864
(24,20)	0.00 %	50.83 %	20.125 dB	7.260 dB	0.970	-7.494	34.166	2.483	3.284	2.109	2.857
(08,55)	0.00 %	51.67 %	15.357 dB	-0.046 dB	0.973	-6.182	38.053	2.170	3.058	2.015	2.576
(12,55)	35.00 %	52.50 %	15.181 dB	0.154 dB	0.974	-6.336	37.197	2.204	3.085	2.028	2.609
(24,55)	0.00 %	52.50 %	16.233 dB	-0.129 dB	0.959	-6.234	37.962	2.175	3.075	2.015	2.588

Table B.19: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

MPDRDL-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	35.00 %	48.33 %	3.688 dB	7.473 dB	0.972	-7.462	37.441	2.641	3.348	2.164	2.960
(12,20)	36.67 %	48.33 %	3.723 dB	7.070 dB	0.913	-7.433	37.860	2.640	3.405	2.163	2.987
(24,20)	35.00 %	48.33 %	3.688 dB	7.428 dB	0.935	-7.409	37.380	2.640	3.386	2.168	2.979
(08,55)	35.83 %	45.83 %	2.380 dB	0.141 dB	0.894	-6.413	40.342	2.276	3.182	2.036	2.686
(12,55)	37.50 %	45.83 %	2.433 dB	0.401 dB	0.921	-6.522	39.569	2.371	3.218	2.079	2.754
(24,55)	37.50 %	45.83 %	2.349 dB	0.131 dB	0.884	-6.418	40.206	2.325	3.223	2.060	2.731

Table B.20: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MPDRVL-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	25.83 %	50.83 %	12.917 dB	-0.286 dB	1.348	-4.825	31.485	1.953	2.600	2.043	2.255
(12,20)	0.00 %	50.83 %	12.688 dB	-0.335 dB	1.375	-4.834	31.250	1.999	2.602	2.066	2.280
(24,20)	0.00 %	50.83 %	12.870 dB	-0.293 dB	1.132	-4.862	31.382	1.966	2.831	2.048	2.377
(08,55)	27.50 %	51.67 %	11.124 dB	-0.538 dB	1.134	-5.498	34.271	2.005	2.826	2.006	2.387
(12,55)	32.50 %	52.50 %	11.038 dB	-0.501 dB	1.145	-5.520	34.328	2.026	2.828	2.014	2.398
(24,55)	30.00 %	52.50 %	11.146 dB	-0.499 dB	0.929	-5.546	36.084	1.980	3.006	1.978	2.460

Table B.21: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

MPDRVL-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	23.33 %	48.33 %	2.094 dB	2.709 dB	1.113	-5.708	34.310	2.157	2.939	2.065	2.520
(12,20)	21.67 %	48.33 %	2.081 dB	2.582 dB	1.235	-5.714	34.213	2.203	2.843	2.088	2.496
(24,20)	17.50 %	48.33 %	2.092 dB	2.693 dB	0.921	-5.738	33.844	2.190	3.162	2.083	2.649
(08,55)	25.00 %	45.83 %	2.172 dB	0.983 dB	0.941	-6.070	36.668	2.184	3.111	2.039	2.613
(12,55)	24.17 %	45.83 %	2.126 dB	1.001 dB	0.848	-6.055	36.589	2.219	3.229	2.057	2.690
(24,55)	20.00 %	45.83 %	2.161 dB	1.034 dB	0.790	-6.086	38.106	2.140	3.227	2.007	2.645

Table B.22: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

GSC (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**78.33 %**	50.83 %	11.148 dB	-2.842 dB	0.808	-5.106	22.816	2.923	3.819	2.550	3.374
(12,20)	**76.67 %**	50.83 %	11.107 dB	-2.829 dB	0.822	-5.057	22.582	2.906	3.796	2.546	3.354
(24,20)	**73.33 %**	50.83 %	11.168 dB	-2.844 dB	1.072	-5.330	24.393	2.779	3.446	2.456	3.112
(08,55)	**77.50 %**	51.67 %	11.866 dB	-3.812 dB	0.822	-4.738	21.997	2.928	3.815	2.581	3.376
(12,55)	**75.00 %**	52.50 %	11.813 dB	-3.778 dB	0.852	-4.845	22.378	2.846	3.731	2.532	3.292
(24,55)	**77.50 %**	52.50 %	11.946 dB	-3.826 dB	1.090	-4.953	23.083	2.903	3.514	2.548	3.211

Table B.23: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

GSC (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**65.83 %**	48.33 %	2.474 dB	-2.474 dB	0.816	-5.465	25.867	2.966	3.810	2.527	3.383
(12,20)	**65.83 %**	48.33 %	2.462 dB	-2.455 dB	0.847	-5.539	25.769	2.927	3.755	2.504	3.336
(24,20)	**70.00 %**	48.33 %	2.474 dB	-2.477 dB	1.128	-5.625	26.233	2.880	3.433	2.473	3.151
(08,55)	**71.67 %**	45.83 %	3.062 dB	-3.635 dB	0.773	-4.950	24.072	3.048	3.919	2.611	3.484
(12,55)	**65.83 %**	45.83 %	3.053 dB	-3.593 dB	0.827	-5.077	25.668	2.994	3.816	2.565	3.401
(24,55)	**70.00 %**	45.83 %	3.063 dB	-3.632 dB	1.008	-5.117	24.913	3.045	3.668	2.593	3.355

Table B.24: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

B.0.2 Results based on Real Data (CPR-Recordings)

Tables of 1^{st} Scenario (Speaker 03 {male} @ MRA=$\{00^\circ\}$)

DS-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**95.00 %**	91.67 %	-0.312 dB	3.583 dB	0.272	1.643	27.533	3.653	4.767	3.291	4.202
(12,20)	**94.17 %**	91.67 %	-0.302 dB	1.189 dB	0.324	1.286	27.021	3.668	4.728	3.279	4.191
(24,20)	**95.83 %**	91.67 %	-0.275 dB	0.647 dB	0.336	1.184	27.023	3.649	4.705	3.264	4.171
(08,55)	**94.17 %**	93.33 %	-0.292 dB	0.825 dB	0.239	0.207	32.766	3.186	4.473	2.941	3.807
(12,55)	**94.17 %**	93.33 %	-0.280 dB	-1.589 dB	0.290	0.193	31.800	3.193	4.434	2.950	3.794
(24,55)	**95.00 %**	93.33 %	-0.202 dB	-2.553 dB	0.356	0.199	30.296	3.247	4.412	2.986	3.813

Table B.25: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

RLSFI-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**95.00 %**	91.67 %	-0.312 dB	3.583 dB	0.272	1.643	27.533	3.653	4.767	3.291	4.202
(12,20)	**94.17 %**	91.67 %	-0.302 dB	1.189 dB	0.324	1.286	27.021	3.668	4.728	3.279	4.191
(24,20)	**95.83 %**	91.67 %	-0.275 dB	0.647 dB	0.336	1.184	27.023	3.649	4.705	3.264	4.171
(08,55)	**94.17 %**	93.33 %	-0.292 dB	0.825 dB	0.239	0.207	32.766	3.186	4.473	2.941	3.807
(12,55)	**94.17 %**	93.33 %	-0.280 dB	-1.589 dB	0.290	0.193	31.800	3.193	4.434	2.950	3.794
(24,55)	**95.00 %**	93.33 %	-0.202 dB	-2.553 dB	0.356	0.199	30.296	3.247	4.412	2.986	3.813

Table B.26: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

MNS-RLSFI-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**95.00 %**	91.67 %	-0.312 dB	3.583 dB	0.272	1.643	27.533	3.653	4.767	3.291	4.202
(12,20)	**94.17 %**	91.67 %	-0.302 dB	1.189 dB	0.324	1.286	27.021	3.668	4.728	3.279	4.191
(24,20)	**95.83 %**	91.67 %	-0.275 dB	0.647 dB	0.336	1.184	27.023	3.649	4.705	3.264	4.171
(08,55)	**94.17 %**	93.33 %	-0.292 dB	0.825 dB	0.239	0.207	32.766	3.186	4.473	2.941	3.807
(12,55)	**94.17 %**	93.33 %	-0.280 dB	-1.589 dB	0.290	0.193	31.800	3.193	4.434	2.950	3.794
(24,55)	**95.00 %**	93.33 %	-0.202 dB	-2.553 dB	0.356	0.199	30.296	3.247	4.412	2.986	3.813

Table B.27: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

MPDRDL-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**95.83 %**	91.67 %	0.592 dB	-3.980 dB	0.427	0.047	34.293	3.163	4.252	2.909	3.682
(12,20)	**94.17 %**	91.67 %	0.609 dB	-7.796 dB	0.495	0.084	33.657	3.176	4.196	2.922	3.662
(24,20)	**92.50 %**	91.67 %	0.361 dB	-8.745 dB	0.501	0.059	33.360	3.128	4.163	2.899	3.622
(08,55)	93.33 %	93.33 %	0.018 dB	-3.269 dB	0.362	-0.306	37.995	2.996	4.185	2.781	3.555
(12,55)	90.00 %	93.33 %	0.079 dB	-6.896 dB	0.471	-0.133	37.304	2.942	4.047	2.771	3.460
(24,55)	92.50 %	93.33 %	-0.093 dB	-7.390 dB	0.498	-0.139	35.380	2.918	4.022	2.773	3.441

Table B.28: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

MPDRVL-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	90.83 %	91.67 %	-0.246 dB	7.932 dB	0.318	1.721	31.152	3.361	4.512	3.131	3.918
(12,20)	91.67 %	91.67 %	-0.278 dB	6.594 dB	0.380	1.576	31.285	3.341	4.435	3.111	3.870
(24,20)	**93.33 %**	91.67 %	-0.287 dB	5.331 dB	0.417	1.491	31.129	3.397	4.432	3.134	3.897
(08,55)	92.50 %	93.33 %	-0.050 dB	8.248 dB	0.310	-0.287	39.658	2.820	4.117	2.686	3.428
(12,55)	92.50 %	93.33 %	-0.338 dB	7.450 dB	0.378	-0.244	39.138	2.845	4.067	2.705	3.417
(24,55)	**94.17 %**	93.33 %	-0.026 dB	5.705 dB	0.396	-0.075	37.403	2.864	4.076	2.736	3.435

Table B.29: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

GSC (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	91.67 %	91.67 %	-0.312 dB	3.583 dB	0.232	-0.710	33.841	3.316	4.549	2.938	3.908
(12,20)	90.00 %	91.67 %	-0.302 dB	1.189 dB	0.228	-0.815	34.002	3.144	4.448	2.848	3.770
(24,20)	85.00 %	91.67 %	-0.275 dB	0.647 dB	0.205	-1.199	36.170	2.787	4.237	2.637	3.479
(08,55)	85.00 %	93.33 %	-0.292 dB	0.825 dB	0.261	-0.493	37.855	2.962	4.270	2.754	3.580
(12,55)	89.17 %	93.33 %	-0.280 dB	-1.589 dB	0.256	-0.663	37.132	3.097	4.363	2.813	3.696
(24,55)	84.17 %	93.33 %	-0.202 dB	-2.553 dB	0.236	-1.202	38.468	2.876	4.238	2.664	3.519

Table B.30: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

Tables of 1^{st} Scenario (Speaker 12 {male} @ MRA=$\{45°\}$)

DS-BF (Speaker 12, MRA = $45°$)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.83 %**	77.50 %	0.548 dB	-4.130 dB	0.241	1.221	22.830	3.814	4.939	3.374	4.381
(12,20)	**82.50 %**	79.17 %	0.680 dB	1.857 dB	0.393	1.128	29.109	3.697	4.656	3.269	4.165
(24,20)	**85.00 %**	77.50 %	0.649 dB	-6.739 dB	0.314	0.880	25.790	3.747	4.797	3.300	4.269
(08,55)	**83.33 %**	79.17 %	-0.019 dB	-6.966 dB	0.252	0.010	28.240	3.391	4.624	3.058	3.997
(12,55)	**83.33 %**	81.67 %	0.105 dB	-0.469 dB	0.508	-0.003	35.096	3.120	4.136	2.880	3.600
(24,55)	**89.17 %**	79.17 %	0.031 dB	-10.010 dB	0.476	0.051	28.244	3.306	4.342	3.020	3.814

Table B.31: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

RLSFI-BF (Speaker 12, MRA = $45°$)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.83 %**	77.50 %	0.548 dB	-4.130 dB	0.241	1.221	22.830	3.814	4.939	3.374	4.381
(12,20)	**82.50 %**	79.17 %	0.680 dB	1.857 dB	0.393	1.128	29.109	3.697	4.656	3.269	4.165
(24,20)	**85.00 %**	77.50 %	0.649 dB	-6.739 dB	0.314	0.880	25.790	3.747	4.797	3.300	4.269
(08,55)	**83.33 %**	79.17 %	-0.019 dB	-6.966 dB	0.252	0.010	28.240	3.391	4.624	3.058	3.997
(12,55)	**83.33 %**	81.67 %	0.105 dB	-0.469 dB	0.508	-0.003	35.096	3.120	4.136	2.880	3.600
(24,55)	**89.17 %**	79.17 %	0.031 dB	-10.010 dB	0.476	0.051	28.244	3.306	4.342	3.020	3.814

Table B.32: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

MNS-RLSFI-BF (Speaker 12, MRA = $45°$)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.83 %**	77.50 %	0.548 dB	-4.130 dB	0.241	1.221	22.830	3.814	4.939	3.374	4.381
(12,20)	**82.50 %**	79.17 %	0.680 dB	1.857 dB	0.393	1.128	29.109	3.697	4.656	3.269	4.165
(24,20)	**85.00 %**	77.50 %	0.649 dB	-6.739 dB	0.314	0.880	25.790	3.747	4.797	3.300	4.269
(08,55)	**83.33 %**	79.17 %	-0.019 dB	-6.966 dB	0.252	0.010	28.240	3.391	4.624	3.058	3.997
(12,55)	**83.33 %**	81.67 %	0.105 dB	-0.469 dB	0.508	-0.003	35.096	3.120	4.136	2.880	3.600
(24,55)	**89.17 %**	79.17 %	0.031 dB	-10.010 dB	0.476	0.051	28.244	3.306	4.342	3.020	3.814

Table B.33: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

MPDRDL-BF (Speaker 12, MRA = $45°$)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.00 %**	77.50 %	0.472 dB	-12.209 dB	0.374	-0.060	33.817	3.199	4.333	2.923	3.741
(12,20)	**82.50 %**	79.17 %	0.384 dB	-7.552 dB	0.600	-0.047	33.372	3.264	4.143	2.957	3.680
(24,20)	74.17 %	77.50 %	0.565 dB	-15.586 dB	0.436	-0.009	32.299	3.214	4.292	2.943	3.732
(08,55)	**80.00 %**	79.17 %	-0.626 dB	-10.413 dB	0.445	-0.248	36.628	3.082	4.164	2.835	3.591
(12,55)	**83.33 %**	81.67 %	-0.205 dB	-4.671 dB	0.748	-0.267	36.259	3.063	3.844	2.827	3.423
(24,55)	**81.67 %**	79.17 %	-0.515 dB	-14.578 dB	0.746	-0.144	32.541	3.004	3.844	2.833	3.403

Table B.34: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

MPDRVL-BF (Speaker 12, MRA = $45°$)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	73.33 %	77.50 %	0.017 dB	1.413 dB	0.293	2.009	24.672	3.625	4.755	3.320	4.189
(12,20)	72.50 %	79.17 %	0.226 dB	8.505 dB	0.447	1.998	30.879	3.577	4.512	3.254	4.029
(24,20)	69.17 %	77.50 %	0.186 dB	-1.038 dB	0.408	1.604	27.289	3.675	4.643	3.300	4.152
(08,55)	72.50 %	79.17 %	-0.464 dB	1.506 dB	0.354	-0.287	33.098	3.128	4.317	2.879	3.699
(12,55)	72.50 %	81.67 %	-0.281 dB	9.404 dB	0.552	-0.440	39.897	2.955	3.948	2.740	3.411
(24,55)	71.67 %	79.17 %	-0.067 dB	-1.609 dB	0.496	-0.003	33.308	3.048	4.120	2.857	3.560

Table B.35: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

GSC (Speaker 12, MRA = $45°$)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	77.50 %	77.50 %	0.548 dB	-4.130 dB	0.241	-0.415	33.094	3.418	4.608	3.010	3.990
(12,20)	76.67 %	79.17 %	0.680 dB	1.857 dB	0.270	-0.484	34.853	3.305	4.494	2.939	3.872
(24,20)	71.67 %	77.50 %	0.649 dB	-6.739 dB	0.224	-0.833	35.930	3.077	4.395	2.801	3.705
(08,55)	77.50 %	79.17 %	-0.019 dB	-6.966 dB	0.307	-0.437	34.325	3.106	4.342	2.851	3.697
(12,55)	75.83 %	81.67 %	0.105 dB	-0.469 dB	0.383	-0.611	36.147	3.111	4.249	2.830	3.649
(24,55)	70.83 %	79.17 %	0.031 dB	-10.010 dB	0.240	-1.082	36.524	2.843	4.232	2.669	3.504

Table B.36: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

Tables of 1^{st} Scenario (Speaker 03 {male} and 12 {male} @ MRA=$\{00^\circ, 45^\circ\}$)

DS-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**61.67 %**	55.00 %	3.522 dB	3.509 dB	0.450	0.237	41.097	2.706	3.892	2.655	3.255
(12,20)	**56.67 %**	55.00 %	3.566 dB	1.084 dB	0.488	0.339	40.327	2.726	3.872	2.676	3.256
(24,20)	**65.83 %**	55.00 %	3.754 dB	0.510 dB	0.489	0.354	39.715	2.723	3.874	2.680	3.258
(08,55)	**62.50 %**	60.83 %	3.004 dB	0.018 dB	0.566	-0.448	44.022	2.641	3.707	2.560	3.122
(12,55)	60.00 %	60.83 %	3.027 dB	-2.638 dB	0.663	-0.240	42.541	2.639	3.619	2.582	3.081
(24,55)	**68.33 %**	60.83 %	3.085 dB	-3.610 dB	0.791	-0.135	38.537	2.688	3.553	2.641	3.083

Table B.37: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

DS-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**37.50 %**	33.33 %	-0.007 dB	-4.647 dB	0.408	-0.298	42.125	2.832	4.001	2.674	3.369
(12,20)	**36.67 %**	35.00 %	-0.003 dB	1.445 dB	0.547	-0.265	43.631	2.803	3.827	2.652	3.265
(24,20)	**43.33 %**	33.33 %	0.027 dB	-7.262 dB	0.450	-0.112	41.245	2.841	3.972	2.696	3.362
(08,55)	**41.67 %**	33.33 %	0.268 dB	-8.363 dB	0.476	-0.761	42.990	2.728	3.861	2.589	3.245
(12,55)	**43.33 %**	30.83 %	0.302 dB	-2.143 dB	0.718	-0.680	45.948	2.626	3.524	2.525	3.019
(24,55)	**50.00 %**	33.33 %	0.227 dB	-11.608 dB	0.704	-0.401	39.153	2.720	3.657	2.635	3.149

Table B.38: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45^\circ$.

RLSFI-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**62.50 %**	55.00 %	4.003 dB	7.984 dB	0.452	-0.641	47.804	2.718	3.836	2.558	3.216
(12,20)	**62.50 %**	55.00 %	3.497 dB	6.328 dB	0.481	-0.444	43.547	2.698	3.833	2.591	3.215
(24,20)	**60.00 %**	55.00 %	2.209 dB	2.249 dB	0.501	-0.514	43.152	2.630	3.775	2.557	3.152
(08,55)	**64.17 %**	60.83 %	4.236 dB	6.054 dB	0.440	-1.117	50.387	2.607	3.760	2.457	3.115
(12,55)	**61.67 %**	60.83 %	4.491 dB	5.683 dB	0.474	-1.181	49.547	2.635	3.748	2.472	3.125
(24,55)	59.17 %	60.83 %	3.929 dB	4.217 dB	0.522	-1.206	50.943	2.482	3.594	2.388	2.968

Table B.39: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

RLSFI-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**41.67 %**	33.33 %	0.643 dB	-10.379 dB	0.371	-0.649	48.972	2.604	3.841	2.495	3.158
(12,20)	34.17 %	35.00 %	0.393 dB	-0.341 dB	0.445	-0.497	49.948	2.563	3.731	2.478	3.080
(24,20)	**38.33 %**	33.33 %	0.472 dB	-8.014 dB	0.443	-0.512	45.633	2.709	3.860	2.577	3.229
(08,55)	**35.83 %**	33.33 %	0.047 dB	-7.156 dB	0.402	-0.848	48.350	2.691	3.866	2.528	3.216
(12,55)	30.00 %	30.83 %	0.222 dB	0.740 dB	0.474	-1.056	51.022	2.530	3.672	2.420	3.031
(24,55)	31.67 %	33.33 %	-0.157 dB	-5.060 dB	0.429	-0.834	52.660	2.547	3.714	2.430	3.056

Table B.40: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45^\circ$.

MNS-RLSFI-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**60.83 %**	55.00 %	3.937 dB	8.002 dB	0.469	-0.669	48.301	2.713	3.812	2.550	3.199
(12,20)	**62.50 %**	55.00 %	3.530 dB	6.322 dB	0.481	-0.430	43.984	2.706	3.833	2.593	3.218
(24,20)	**60.00 %**	55.00 %	2.262 dB	2.256 dB	0.500	-0.511	43.535	2.627	3.771	2.553	3.148
(08,55)	57.50 %	60.83 %	4.008 dB	6.062 dB	0.474	-1.134	50.882	2.593	3.711	2.446	3.083
(12,55)	**61.67 %**	60.83 %	4.479 dB	5.683 dB	0.483	-1.186	49.231	2.635	3.742	2.474	3.123
(24,55)	59.17 %	60.83 %	3.931 dB	4.214 dB	0.509	-1.205	50.842	2.486	3.611	2.391	2.979

Table B.41: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0^\circ$.

MNS-RLSFI-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**36.67 %**	33.33 %	0.582 dB	-10.478 dB	0.397	-0.635	48.731	2.606	3.817	2.498	3.147
(12,20)	**39.17 %**	35.00 %	0.399 dB	-0.378 dB	0.436	-0.472	49.697	2.613	3.772	2.505	3.126
(24,20)	**40.00 %**	33.33 %	0.503 dB	-7.954 dB	0.444	-0.510	45.588	2.736	3.876	2.590	3.250
(08,55)	**35.83 %**	33.33 %	0.055 dB	-7.123 dB	0.472	-0.858	48.823	2.660	3.772	2.510	3.152
(12,55)	30.00 %	30.83 %	0.222 dB	0.761 dB	0.481	-1.057	50.657	2.535	3.671	2.425	3.034
(24,55)	32.50 %	33.33 %	-0.150 dB	-5.042 dB	0.431	-0.832	52.496	2.558	3.719	2.437	3.065

Table B.42: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45^\circ$.

MPDRDL-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**61.67 %**	55.00 %	1.527 dB	-4.655 dB	0.663	-0.349	45.740	2.559	3.542	2.515	2.995
(12,20)	**61.67 %**	55.00 %	1.344 dB	-8.837 dB	0.731	-0.122	44.074	2.558	3.486	2.540	2.970
(24,20)	**61.67 %**	55.00 %	1.638 dB	-8.987 dB	0.723	-0.148	43.717	2.510	3.469	2.519	2.939
(08,55)	**64.17 %**	60.83 %	2.413 dB	-3.351 dB	0.805	-0.612	46.598	2.477	3.338	2.453	2.849
(12,55)	59.17 %	60.83 %	2.988 dB	-6.789 dB	0.925	-0.324	44.731	2.492	3.241	2.492	2.813
(24,55)	53.33 %	60.83 %	3.780 dB	-7.127 dB	1.078	-0.323	41.976	2.454	3.086	2.493	2.723

Table B.43: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0°$.

MPDRDL-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	27.50 %	33.33 %	1.081 dB	-12.763 dB	0.551	-0.481	46.625	2.605	3.677	2.523	3.083
(12,20)	30.00 %	35.00 %	0.882 dB	-7.633 dB	0.754	-0.387	45.018	2.596	3.477	2.535	2.982
(24,20)	30.83 %	33.33 %	1.082 dB	-16.125 dB	0.584	-0.265	43.772	2.619	3.678	2.563	3.097
(08,55)	30.00 %	33.33 %	0.347 dB	-11.826 dB	0.660	-0.619	47.679	2.526	3.508	2.469	2.956
(12,55)	**38.33 %**	30.83 %	0.378 dB	-5.984 dB	0.926	-0.584	45.604	2.506	3.241	2.476	2.818
(24,55)	26.67 %	33.33 %	0.227 dB	-15.344 dB	0.978	-0.382	41.629	2.495	3.217	2.511	2.810

Table B.44: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

MPDRVL-BF (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	50.00 %	55.00 %	2.545 dB	8.265 dB	0.521	-0.544	46.143	2.427	3.605	2.437	2.958
(12,20)	**56.67 %**	55.00 %	3.144 dB	7.192 dB	0.594	-0.289	46.140	2.400	3.513	2.440	2.899
(24,20)	**56.67 %**	55.00 %	4.199 dB	5.892 dB	0.604	-0.099	45.021	2.502	3.575	2.509	2.984
(08,55)	55.00 %	60.83 %	3.070 dB	9.582 dB	0.681	-1.786	51.427	2.248	3.285	2.236	2.695
(12,55)	55.00 %	60.83 %	5.078 dB	9.218 dB	0.811	-1.543	50.493	2.275	3.176	2.271	2.657
(24,55)	53.33 %	60.83 %	6.118 dB	7.233 dB	0.878	-1.127	46.865	2.332	3.174	2.349	2.693

Table B.45: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0°$.

MPDRVL-BF (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	24.17 %	33.33 %	-0.024 dB	1.209 dB	0.515	-0.944	45.786	2.644	3.745	2.518	3.138
(12,20)	25.00 %	35.00 %	0.023 dB	8.391 dB	0.667	-1.083	47.383	2.668	3.589	2.509	3.068
(24,20)	24.17 %	33.33 %	0.055 dB	-1.378 dB	0.600	-0.418	44.627	2.763	3.740	2.616	3.199
(08,55)	22.50 %	33.33 %	-0.116 dB	1.702 dB	0.626	-2.108	49.662	2.388	3.441	2.295	2.848
(12,55)	25.83 %	30.83 %	-0.227 dB	9.851 dB	0.845	-2.344	52.235	2.314	3.149	2.227	2.659
(24,55)	22.50 %	33.33 %	0.032 dB	-0.770 dB	0.813	-1.435	46.453	2.390	3.280	2.361	2.776

Table B.46: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

GSC (Speaker 03, MRA = 0°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**60.00 %**	55.00 %	3.522 dB	3.509 dB	0.406	-1.189	43.852	2.626	3.864	2.507	3.193
(12,20)	**56.67 %**	55.00 %	3.566 dB	1.084 dB	0.385	-1.131	43.003	2.538	3.840	2.475	3.139
(24,20)	**64.17 %**	55.00 %	3.754 dB	0.510 dB	0.379	-1.649	44.148	2.200	3.633	2.273	2.862
(08,55)	58.33 %	60.83 %	3.004 dB	0.018 dB	0.545	-0.767	45.651	2.536	3.650	2.478	3.036
(12,55)	57.50 %	60.83 %	3.027 dB	-2.638 dB	0.480	-0.833	44.962	2.590	3.756	2.505	3.119
(24,55)	52.50 %	60.83 %	3.085 dB	-3.610 dB	0.449	-1.354	46.309	2.295	3.598	2.322	2.888

Table B.47: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 0°$.

GSC (Speaker 12, MRA = 45°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	32.50 %	33.33 %	-0.007 dB	-4.647 dB	0.390	-1.087	45.182	2.755	3.946	2.566	3.295
(12,20)	30.83 %	35.00 %	-0.003 dB	1.445 dB	0.407	-1.197	45.128	2.699	3.896	2.533	3.243
(24,20)	28.33 %	33.33 %	0.027 dB	-7.262 dB	0.353	-1.506	46.260	2.438	3.783	2.381	3.052
(08,55)	**37.50 %**	33.33 %	0.268 dB	-8.363 dB	0.540	-0.891	44.591	2.629	3.721	2.523	3.122
(12,55)	**32.50 %**	30.83 %	0.302 dB	-2.143 dB	0.596	-1.123	45.629	2.603	3.638	2.488	3.064
(24,55)	**36.67 %**	33.33 %	0.227 dB	-11.608 dB	0.404	-1.614	45.550	2.346	3.682	2.335	2.957

Table B.48: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 45°$.

Tables of 2^{nd} Scenario (Speaker 03 {male} @ MRA=$\{300^\circ\}$)

DS-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**95.00 %**	91.67 %	-0.264 dB	-1.246 dB	0.224	1.465	27.499	3.585	4.777	3.247	4.173
(12,20)	**95.00 %**	93.33 %	-0.254 dB	-2.132 dB	0.252	1.190	26.002	3.575	4.755	3.236	4.160
(24,20)	**95.83 %**	93.33 %	-0.309 dB	-2.499 dB	0.279	1.164	25.177	3.623	4.763	3.263	4.191
(08,55)	93.33 %	93.33 %	-0.550 dB	-4.317 dB	0.239	0.030	34.998	3.121	4.414	2.883	3.739
(12,55)	95.00 %	95.00 %	-0.753 dB	-5.196 dB	0.289	-0.034	32.407	3.184	4.424	2.927	3.782
(24,55)	93.33 %	95.00 %	-0.958 dB	-5.819 dB	0.359	0.023	30.810	3.129	4.332	2.915	3.713

Table B.49: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300^\circ$.

RLSFI-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**95.00 %**	91.67 %	-0.264 dB	-1.246 dB	0.224	1.465	27.499	3.585	4.777	3.247	4.173
(12,20)	**95.00 %**	93.33 %	-0.254 dB	-2.132 dB	0.252	1.190	26.002	3.575	4.755	3.236	4.160
(24,20)	**95.83 %**	93.33 %	-0.309 dB	-2.499 dB	0.279	1.164	25.177	3.623	4.763	3.263	4.191
(08,55)	93.33 %	93.33 %	-0.550 dB	-4.317 dB	0.239	0.030	34.998	3.121	4.414	2.883	3.739
(12,55)	95.00 %	95.00 %	-0.753 dB	-5.196 dB	0.289	-0.034	32.407	3.184	4.424	2.927	3.782
(24,55)	93.33 %	95.00 %	-0.958 dB	-5.819 dB	0.359	0.023	30.810	3.129	4.332	2.915	3.713

Table B.50: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300^\circ$.

MNS-RLSFI-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**95.00 %**	91.67 %	-0.264 dB	-1.246 dB	0.224	1.465	27.499	3.585	4.777	3.247	4.173
(12,20)	**95.00 %**	93.33 %	-0.254 dB	-2.132 dB	0.252	1.190	26.002	3.575	4.755	3.236	4.160
(24,20)	**95.83 %**	93.33 %	-0.309 dB	-2.499 dB	0.279	1.164	25.177	3.623	4.763	3.263	4.191
(08,55)	93.33 %	93.33 %	-0.550 dB	-4.317 dB	0.239	0.030	34.998	3.121	4.414	2.883	3.739
(12,55)	95.00 %	95.00 %	-0.753 dB	-5.196 dB	0.289	-0.034	32.407	3.184	4.424	2.927	3.782
(24,55)	93.33 %	95.00 %	-0.958 dB	-5.819 dB	0.359	0.023	30.810	3.129	4.332	2.915	3.713

Table B.51: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300^\circ$.

MPDRDL-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**92.50 %**	91.67 %	0.650 dB	-7.566 dB	0.314	0.299	35.050	3.010	4.269	2.846	3.611
(12,20)	90.83 %	93.33 %	0.686 dB	-10.590 dB	0.390	0.192	35.380	2.996	4.179	2.830	3.558
(24,20)	91.67 %	93.33 %	0.773 dB	-10.675 dB	0.402	0.215	34.169	2.960	4.157	2.823	3.532
(08,55)	93.33 %	93.33 %	-0.639 dB	-9.736 dB	0.361	-0.011	38.756	2.870	4.104	2.734	3.449
(12,55)	93.33 %	95.00 %	-0.712 dB	-12.394 dB	0.488	-0.014	36.760	2.853	3.980	2.739	3.383
(24,55)	89.17 %	95.00 %	-0.882 dB	-11.690 dB	0.672	-0.017	34.570	2.818	3.790	2.738	3.277

Table B.52: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300^\circ$.

MPDRVL-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**92.50 %**	91.67 %	-0.200 dB	3.262 dB	0.266	1.126	32.529	3.233	4.476	3.023	3.833
(12,20)	91.67 %	93.33 %	-0.204 dB	3.912 dB	0.329	1.082	31.053	3.244	4.431	3.036	3.820
(24,20)	90.00 %	93.33 %	-0.342 dB	2.050 dB	0.338	1.315	28.997	3.324	4.489	3.103	3.894
(08,55)	92.50 %	93.33 %	-0.002 dB	4.237 dB	0.296	-0.487	40.886	2.801	4.109	2.656	3.411
(12,55)	92.50 %	95.00 %	-0.324 dB	4.909 dB	0.372	-0.521	38.716	2.831	4.068	2.683	3.411
(24,55)	91.67 %	95.00 %	-0.567 dB	3.125 dB	0.435	-0.149	36.285	2.839	4.031	2.728	3.402

Table B.53: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300^\circ$.

GSC (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	88.33 %	91.67 %	-0.264 dB	-1.246 dB	0.182	-0.817	33.528	3.350	4.625	2.949	3.963
(12,20)	87.50 %	93.33 %	-0.254 dB	-2.132 dB	0.207	-1.063	35.217	3.164	4.471	2.833	3.789
(24,20)	85.83 %	93.33 %	-0.309 dB	-2.499 dB	0.193	-1.652	35.706	2.922	4.335	2.677	3.597
(08,55)	88.33 %	93.33 %	-0.550 dB	-4.317 dB	0.209	-0.691	37.763	2.972	4.330	2.747	3.615
(12,55)	89.17 %	95.00 %	-0.753 dB	-5.196 dB	0.231	-0.906	37.769	2.987	4.317	2.740	3.616
(24,55)	85.83 %	95.00 %	-0.958 dB	-5.819 dB	0.207	-1.492	37.444	2.844	4.258	2.637	3.515

Table B.54: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300^\circ$.

Tables of 2^{nd} Scenario (Speaker 12 {male} @ MRA={120°})

DS-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.00 %**	74.17 %	0.084 dB	2.080 dB	0.187	1.416	24.999	3.778	4.954	3.354	4.365
(12,20)	79.17 %	79.17 %	0.352 dB	3.950 dB	0.400	0.990	29.768	3.703	4.646	3.258	4.162
(24,20)	**80.00 %**	79.17 %	0.312 dB	3.874 dB	0.398	1.054	29.644	3.752	4.679	3.286	4.203
(08,55)	79.17 %	80.83 %	0.290 dB	-0.991 dB	0.196	0.042	31.661	3.236	4.558	2.962	3.877
(12,55)	**82.50 %**	79.17 %	0.562 dB	0.886 dB	0.551	-0.116	33.205	3.258	4.192	2.952	3.702
(24,55)	**88.33 %**	79.17 %	0.382 dB	0.614 dB	0.599	-0.021	31.552	3.234	4.143	2.958	3.670

Table B.55: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

RLSFI-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.00 %**	74.17 %	0.084 dB	2.080 dB	0.187	1.416	24.999	3.778	4.954	3.354	4.365
(12,20)	79.17 %	79.17 %	0.352 dB	3.950 dB	0.400	0.990	29.768	3.703	4.646	3.258	4.162
(24,20)	**80.00 %**	79.17 %	0.312 dB	3.874 dB	0.398	1.054	29.644	3.752	4.679	3.286	4.203
(08,55)	79.17 %	80.83 %	0.290 dB	-0.991 dB	0.196	0.042	31.661	3.236	4.558	2.962	3.877
(12,55)	**82.50 %**	79.17 %	0.562 dB	0.886 dB	0.551	-0.116	33.205	3.258	4.192	2.952	3.702
(24,55)	**88.33 %**	79.17 %	0.382 dB	0.614 dB	0.599	-0.021	31.552	3.234	4.143	2.958	3.670

Table B.56: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MNS-RLSFI-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**80.00 %**	74.17 %	0.084 dB	2.080 dB	0.187	1.416	24.999	3.778	4.954	3.354	4.365
(12,20)	79.17 %	79.17 %	0.352 dB	3.950 dB	0.400	0.990	29.768	3.703	4.646	3.258	4.162
(24,20)	**80.00 %**	79.17 %	0.312 dB	3.874 dB	0.398	1.054	29.644	3.752	4.679	3.286	4.203
(08,55)	79.17 %	80.83 %	0.290 dB	-0.991 dB	0.196	0.042	31.661	3.236	4.558	2.962	3.877
(12,55)	**82.50 %**	79.17 %	0.562 dB	0.886 dB	0.551	-0.116	33.205	3.258	4.192	2.952	3.702
(24,55)	**88.33 %**	79.17 %	0.382 dB	0.614 dB	0.599	-0.021	31.552	3.234	4.143	2.958	3.670

Table B.57: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MPDRDL-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**78.33 %**	74.17 %	-0.770 dB	-6.902 dB	0.292	-0.025	34.641	3.220	4.423	2.929	3.794
(12,20)	79.17 %	79.17 %	-0.232 dB	-6.020 dB	0.642	-0.045	34.542	3.213	4.059	2.925	3.610
(24,20)	79.17 %	79.17 %	-0.340 dB	-5.983 dB	0.673	-0.008	35.046	3.117	3.965	2.878	3.514
(08,55)	80.83 %	80.83 %	0.299 dB	-4.920 dB	0.302	-0.215	37.444	2.987	4.246	2.786	3.582
(12,55)	**83.33 %**	79.17 %	1.140 dB	-4.012 dB	0.800	-0.136	36.195	2.939	3.717	2.777	3.297
(24,55)	**83.33 %**	79.17 %	1.118 dB	-4.124 dB	1.082	-0.137	33.562	2.873	3.410	2.764	3.118

Table B.58: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MPDRVL-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	74.17 %	74.17 %	-0.594 dB	7.985 dB	0.235	2.614	27.260	3.540	4.741	3.300	4.133
(12,20)	74.17 %	79.17 %	-0.178 dB	11.559 dB	0.512	2.254	32.082	3.560	4.424	3.253	3.973
(24,20)	74.17 %	79.17 %	0.170 dB	9.543 dB	0.499	2.243	31.524	3.577	4.453	3.264	3.997
(08,55)	71.67 %	80.83 %	-0.669 dB	9.020 dB	0.262	-0.089	35.497	2.985	4.304	2.807	3.614
(12,55)	70.00 %	79.17 %	-0.277 dB	13.038 dB	0.628	-0.450	36.708	2.986	3.917	2.776	3.419
(24,55)	69.17 %	79.17 %	-0.186 dB	10.686 dB	0.712	0.106	35.207	2.910	3.799	2.785	3.326

Table B.59: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

GSC (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**75.83 %**	74.17 %	0.084 dB	2.080 dB	0.199	-0.619	34.417	3.399	4.628	2.979	3.987
(12,20)	75.00 %	79.17 %	0.352 dB	3.950 dB	0.256	-0.391	35.938	3.123	4.390	2.851	3.726
(24,20)	70.83 %	79.17 %	0.312 dB	3.874 dB	0.229	-0.798	37.214	2.850	4.241	2.685	3.510
(08,55)	70.00 %	80.83 %	0.290 dB	-0.991 dB	0.218	-0.676	35.841	3.024	4.370	2.786	3.666
(12,55)	71.67 %	79.17 %	0.562 dB	0.886 dB	0.337	-0.593	36.330	3.103	4.290	2.826	3.665
(24,55)	60.83 %	79.17 %	0.382 dB	0.614 dB	0.252	-1.036	37.818	2.451	3.972	2.476	3.174

Table B.60: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

Tables of 2^{nd} Scenario (Speaker 03 {male} and 12 {male} @ MRA=$\{300°, 120°\}$)

DS-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**61.67 %**	50.83 %	1.814 dB	-1.981 dB	0.375	0.041	44.421	2.725	3.951	2.628	3.285
(12,20)	**60.83 %**	51.67 %	1.831 dB	-2.935 dB	0.389	0.109	42.008	2.712	3.950	2.643	3.284
(24,20)	**61.67 %**	51.67 %	1.858 dB	-3.234 dB	0.390	0.163	40.832	2.754	3.984	2.675	3.325
(08,55)	**61.67 %**	55.83 %	2.098 dB	-5.366 dB	0.447	-0.586	45.927	2.568	3.769	2.503	3.111
(12,55)	**60.00 %**	50.83 %	2.234 dB	-6.033 dB	0.511	-0.495	42.760	2.593	3.746	2.543	3.120
(24,55)	**66.67 %**	50.83 %	2.342 dB	-7.133 dB	0.571	-0.330	39.328	2.634	3.739	2.597	3.146

Table B.61: This table contains the resulting measurements of the enhanced file of speaker 03 with $m\phi_s = 300°$.

DS-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**38.33 %**	30.83 %	0.444 dB	1.314 dB	0.324	-1.333	42.814	2.839	4.086	2.607	3.414
(12,20)	37.50 %	37.50 %	0.493 dB	3.117 dB	0.500	-0.805	44.476	2.809	3.872	2.615	3.288
(24,20)	**42.50 %**	37.50 %	0.377 dB	3.080 dB	0.476	-0.768	43.400	2.842	3.927	2.640	3.335
(08,55)	**35.83 %**	33.33 %	1.541 dB	-1.914 dB	0.361	-1.182	43.222	2.692	3.956	2.544	3.274
(12,55)	**37.50 %**	35.83 %	1.101 dB	0.336 dB	0.757	-0.773	44.321	2.655	3.516	2.544	3.034
(24,55)	**49.17 %**	35.83 %	1.524 dB	-0.604 dB	0.791	-0.625	40.947	2.727	3.555	2.611	3.097

Table B.62: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

RLSFI-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**61.67 %**	50.83 %	-0.739 dB	-8.654 dB	0.440	-0.684	52.412	2.371	3.598	2.357	2.910
(12,20)	**52.50 %**	51.67 %	0.526 dB	-4.146 dB	0.380	-1.405	48.716	2.486	3.762	2.393	3.059
(24,20)	46.67 %	51.67 %	0.881 dB	-5.386 dB	0.462	-0.680	50.281	2.472	3.655	2.421	2.995
(08,55)	55.83 %	55.83 %	3.603 dB	-1.048 dB	0.389	-0.799	51.991	2.627	3.808	2.475	3.145
(12,55)	50.83 %	50.83 %	1.472 dB	-4.528 dB	0.398	-0.591	54.053	2.323	3.598	2.329	2.882
(24,55)	46.67 %	50.83 %	2.395 dB	0.694 dB	0.411	-0.859	55.447	2.458	3.653	2.367	2.974

Table B.63: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

RLSFI-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**32.50 %**	30.83 %	-0.130 dB	-3.097 dB	0.364	-1.334	46.782	2.675	3.911	2.501	3.234
(12,20)	30.83 %	37.50 %	-0.480 dB	3.515 dB	0.466	-1.341	47.842	2.694	3.807	2.502	3.189
(24,20)	30.00 %	37.50 %	0.243 dB	2.252 dB	0.562	-1.082	48.443	2.783	3.757	2.557	3.207
(08,55)	**36.67 %**	33.33 %	1.132 dB	2.301 dB	0.357	-1.275	51.536	2.675	3.875	2.472	3.204
(12,55)	30.83 %	35.83 %	0.582 dB	2.679 dB	0.430	-0.590	52.237	2.533	3.707	2.442	3.047
(24,55)	30.00 %	35.83 %	0.469 dB	7.562 dB	0.458	-1.324	53.780	2.611	3.712	2.422	3.085

Table B.64: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MNS-RLSFI-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**61.67 %**	50.83 %	-0.984 dB	-8.508 dB	0.468	-0.723	52.477	2.360	3.562	2.349	2.887
(12,20)	51.67 %	51.67 %	0.536 dB	-4.140 dB	0.369	-1.405	48.672	2.488	3.776	2.394	3.067
(24,20)	46.67 %	51.67 %	0.887 dB	-5.387 dB	0.465	-0.678	50.657	2.468	3.647	2.417	2.989
(08,55)	50.00 %	55.83 %	3.349 dB	-1.000 dB	0.414	-0.844	53.131	2.603	3.759	2.453	3.106
(12,55)	**53.33 %**	50.83 %	1.288 dB	-4.457 dB	0.398	-0.616	53.899	2.290	3.579	2.312	2.856
(24,55)	47.50 %	50.83 %	2.430 dB	0.690 dB	0.407	-0.853	55.495	2.455	3.655	2.365	2.974

Table B.65: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

MNS-RLSFI-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**31.67 %**	30.83 %	-0.095 dB	-2.963 dB	0.392	-1.390	46.975	2.668	3.876	2.493	3.212
(12,20)	29.17 %	37.50 %	-0.483 dB	3.518 dB	0.457	-1.339	47.953	2.692	3.815	2.501	3.192
(24,20)	30.00 %	37.50 %	0.243 dB	2.238 dB	0.565	-1.073	48.443	2.781	3.753	2.557	3.204
(08,55)	**35.83 %**	33.33 %	1.138 dB	2.344 dB	0.376	-1.314	51.752	2.668	3.849	2.464	3.187
(12,55)	30.83 %	35.83 %	0.606 dB	2.733 dB	0.450	-0.608	52.472	2.512	3.672	2.429	3.018
(24,55)	30.00 %	35.83 %	0.469 dB	7.563 dB	0.456	-1.323	53.607	2.614	3.717	2.425	3.089

Table B.66: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MPDRDL-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	45.83 %	50.83 %	1.509 dB	-9.472 dB	0.543	-0.154	47.310	2.376	3.541	2.429	2.897
(12,20)	45.83 %	51.67 %	1.053 dB	-11.988 dB	0.620	-0.101	46.252	2.358	3.461	2.431	2.852
(24,20)	46.67 %	51.67 %	1.968 dB	-11.241 dB	0.635	-0.114	45.491	2.289	3.410	2.403	2.793
(08,55)	47.50 %	55.83 %	1.577 dB	-8.881 dB	0.626	-0.394	49.414	2.255	3.365	2.341	2.743
(12,55)	48.33 %	50.83 %	1.383 dB	-11.067 dB	0.745	-0.258	45.923	2.231	3.258	2.363	2.687
(24,55)	44.17 %	50.83 %	2.641 dB	-10.238 dB	0.979	-0.288	44.078	2.219	3.027	2.368	2.570

Table B.67: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

MPDRDL-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	24.17 %	30.83 %	-0.635 dB	-5.727 dB	0.470	-0.838	49.368	2.403	3.615	2.384	2.943
(12,20)	23.33 %	37.50 %	-0.494 dB	-4.913 dB	0.819	-0.476	48.409	2.366	3.241	2.396	2.740
(24,20)	20.83 %	37.50 %	-0.584 dB	-4.277 dB	0.843	-0.501	48.501	2.298	3.174	2.361	2.672
(08,55)	22.50 %	33.33 %	0.152 dB	-4.656 dB	0.486	-0.762	48.025	2.435	3.629	2.414	2.970
(12,55)	20.00 %	35.83 %	0.231 dB	-3.538 dB	0.992	-0.485	46.975	2.376	3.082	2.410	2.670
(24,55)	16.67 %	35.83 %	-0.288 dB	-2.950 dB	1.264	-0.521	44.214	2.260	2.757	2.372	2.457

Table B.68: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

MPDRVL-BF (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	50.83 %	50.83 %	1.781 dB	3.729 dB	0.476	-1.201	50.267	2.363	3.576	2.336	2.901
(12,20)	49.17 %	51.67 %	2.964 dB	4.608 dB	0.532	-1.057	48.245	2.386	3.550	2.370	2.905
(24,20)	45.83 %	51.67 %	3.144 dB	2.897 dB	0.535	-0.603	46.516	2.456	3.605	2.445	2.972
(08,55)	0.00 %	55.83 %	1.861 dB	5.602 dB	0.560	-2.110	52.550	2.146	3.338	2.159	2.667
(12,55)	50.00 %	50.83 %	3.083 dB	6.127 dB	0.649	-1.908	50.062	2.207	3.305	2.218	2.688
(24,55)	**52.50 %**	50.83 %	3.614 dB	4.847 dB	0.738	-1.482	47.210	2.194	3.232	2.259	2.652

Table B.69: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

MPDRVL-BF (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	20.00 %	30.83 %	-0.411 dB	8.105 dB	0.438	-2.475	46.373	2.623	3.807	2.407	3.157
(12,20)	20.00 %	37.50 %	-0.335 dB	11.746 dB	0.664	-2.123	48.745	2.588	3.532	2.396	2.996
(24,20)	18.33 %	37.50 %	-0.184 dB	9.751 dB	0.655	-1.703	47.764	2.634	3.578	2.452	3.045
(08,55)	22.50 %	33.33 %	0.124 dB	9.315 dB	0.477	-3.040	48.823	2.422	3.623	2.259	2.958
(12,55)	23.33 %	35.83 %	0.184 dB	13.070 dB	0.879	-2.747	49.048	2.464	3.233	2.295	2.784
(24,55)	17.50 %	35.83 %	-0.104 dB	11.494 dB	0.984	-2.375	46.765	2.371	3.089	2.291	2.672

Table B.70: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

GSC (Speaker 03, MRA = 300°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**57.50 %**	50.83 %	1.814 dB	-1.981 dB	0.360	-1.225	45.416	2.652	3.913	2.507	3.227
(12,20)	**56.67 %**	51.67 %	1.831 dB	-2.935 dB	0.382	-1.400	45.046	2.563	3.839	2.455	3.146
(24,20)	**55.00 %**	51.67 %	1.858 dB	-3.234 dB	0.374	-2.093	45.627	2.301	3.685	2.283	2.936
(08,55)	53.33 %	55.83 %	2.098 dB	-5.366 dB	0.433	-0.949	46.847	2.498	3.733	2.440	3.056
(12,55)	**51.67 %**	50.83 %	2.234 dB	-6.033 dB	0.463	-1.154	45.607	2.497	3.712	2.436	3.048
(24,55)	**54.17 %**	50.83 %	2.342 dB	-7.133 dB	0.392	-1.837	45.994	2.327	3.679	2.308	2.944

Table B.71: This table contains the resulting measurements of the enhanced file of speaker 03 with $\phi_s = 300°$.

GSC (Speaker 12, MRA = 120°)

	WRe	WRn	iGSINR	GAIN	LLR	sSNR	WSS	PESQ	Csig	Cbak	Covrl
(08,20)	**34.17 %**	30.83 %	0.444 dB	1.314 dB	0.337	-1.501	45.345	2.752	3.998	2.538	3.320
(12,20)	32.50 %	37.50 %	0.493 dB	3.117 dB	0.395	-1.131	45.385	2.629	3.863	2.502	3.190
(24,20)	33.33 %	37.50 %	0.377 dB	3.080 dB	0.378	-1.658	45.603	2.475	3.786	2.393	3.073
(08,55)	32.50 %	33.33 %	1.541 dB	-1.914 dB	0.380	-1.220	43.675	2.681	3.925	2.533	3.252
(12,55)	31.67 %	35.83 %	1.101 dB	0.336 dB	0.572	-1.029	44.150	2.633	3.694	2.519	3.111
(24,55)	**37.50 %**	35.83 %	1.524 dB	-0.604 dB	0.424	-1.464	44.967	2.313	3.646	2.332	2.924

Table B.72: This table contains the resulting measurements of the enhanced file of speaker 12 with $\phi_s = 120°$.

C
Abbreviations

ADC	Analog to Digital Converter
BF	Beamformer
BW	Beamwidth
CHiME	Computational Hearing in Multisource Environments
CPR	Cocktail Party Room
CPU	Central Processing Unit
Cbak	Composite Measure for Background Noise Distortion
C-BAK	Composite Measure for Background Noise Distortion
Corvl	Composite Measure for Overall Quality
C-OVRL	Composite Measure for Overall Quality
Csig	Composite Measure for Signal Distortion
C-SIG	Composite Measure for Signal Distortion
DAC	Digital to Analog Converter
DI	Directivity Index
DI-BF	Data-Independent Beamformer
DD-BF	Data-Dependent Beamformer
DIRHA	Distant Speech Interaction for Robust Home Applications
DOA	Direction of Arrival
DS	Delay-&-Sum
GSC	Generalized Sidelobe Canceller
LLR	Log-Likelihood Ratio
LMS	Least Mean Square
MMSE	Minimum Mean Square Error
MNS-RLSFI	Multiple Null Synthesis Robust Least Squares Frequency Invariant
MPDR	Minimum Power Distortionless Response
MPDRDL	Minimum Power Distortionless Response with Diagonal Loading
MPDRVL	Minimum Power Distortionless Response with Variable Loading
MSR	Main to Side Lobe Ratio
MVDR	Minimum Variance Distortionless Response
NLMS	Normalized Least Mean Square
OLA	Overlap and Add
PESQ	Perceptual Evaluation of Speech Quality

RLSFI	Robust Least Squares Frequency Invariant
RMS	Root Mean Square
SNR	Signal to Noise Ratio
sSNR	segmental Signal to Noise Ratio
TDOA	Time Delay of Arrival
UCA	Uniform Circular Array
ULA	Uniform Linear Array
VAD	Voice Activity Detection
WRe	Word Recognition Rate of the enhanced signal
WRn	Word Recognition Rate of the signal captured by the nearest microphone
WSS	Weighted Spectral Slope Difference

D Symbols

c	sound velocity
t	time variable
T_0	time interval
f	frequency variable
f_s	sampling frequency
f_{sa}	spatial aliasing frequency
f_{gl}	grating lobe frequency
λ_{gl}	grating lobe wave length
ω	angular frequency
ω_0	angular frequency of a propagating plane wave
d	distance between each microphone in case of a ULA or diameter of a UCA
τ_n	delay of channel n
$\boldsymbol{r}_n$	microphone position vector
$\boldsymbol{k}$	wavevector
$\boldsymbol{k}_0$	direction of propagation of a plane wave
A_0	amplitude of a plane wave
$y(t)$	(mono-)output signal
$s(\boldsymbol{r}_n, t)$	signal captured by microphone n
$p(\boldsymbol{r}, t)$	acoustic pressure at position $\boldsymbol{r}$ in spatio-temporal domain
$P(\boldsymbol{k}, w)$	acoustic pressure at position $\boldsymbol{r}$ in wavevector-frequency domain
θ	elevation
ϕ	azimuth
ϕ_s	azimuth angle of the desired source / steering direction
ϕ_n	azimuth angle of microphone n
ϕ_l	azimuth angle of looking direction
ϕ_c	azimuth angle of competing source
ϕ_{ML1}	azimuth angle of the main lobe of a ULA
ϕ_{ML2}	azimuth angle of the main lobe of a ULA due to front-back ambiguity
θ_n	elevation angle of microphone n
θ_s	elevation angle of the desired source
$d\phi$	angle resolution

$h(\boldsymbol{r}, t)$	spatio-temporal impulse response
$H(\boldsymbol{k}, \omega)$	wavevector-frequency transfer function
$H(\omega, \varphi, \phi_s)$	beam pattern
$D_n(\omega)$	sound capture model of microphone n in frequency domain
$B_n(\omega)$	beamformer kernel
$N(\omega)$	noise source in frequency domain
$R(\omega)$	sound source in frequency domain
N	number of microphones
$U(f)$	real microphone characteristic in frequency domain
$U_{opt}(f)$	ideal microphone characteristic in frequency domain
$M(f)$	variations in sensitivity of a microphone in frequency domain
$\varphi(f)$	phase deviations in frequency domain
G_m	individual gain of channel m
$\bar{L}$	averaged RMS over all channels
L_m	RMS of channel m
$\tilde{L}(d)$	distance-dependent level interpolation
SL_{num}	number of side lobes