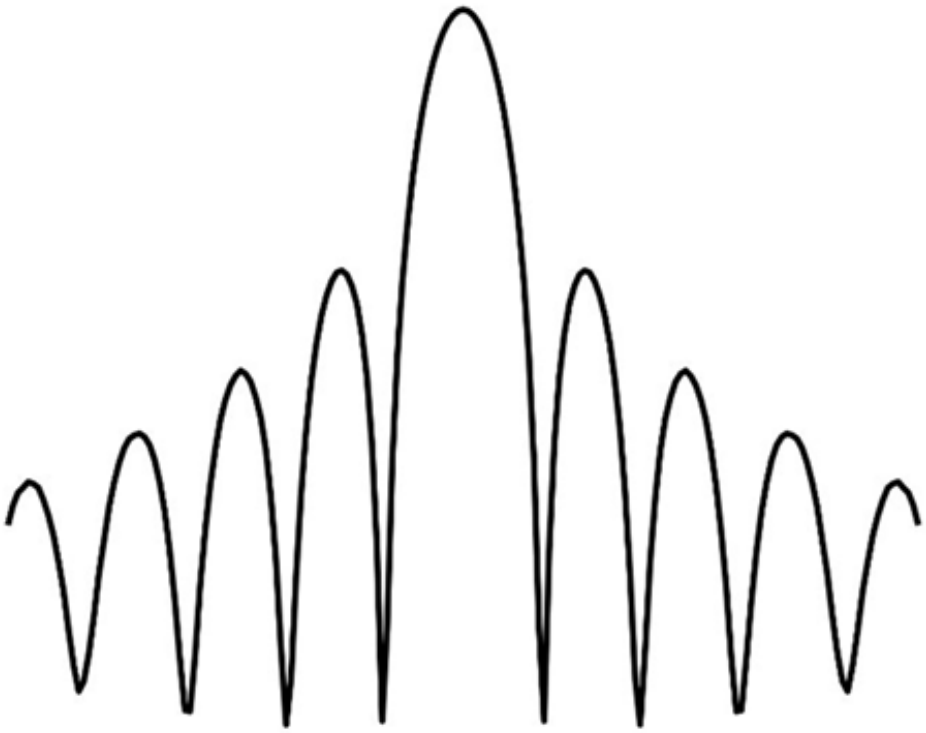


Acoustic Source Localization for an Indoor Pass-By Measurement System

A Beamforming Approach Using a One-Dimensional Sparse Microphone Array

Master's Thesis in Sound and Vibration

Gustav Seger
Semir Sehic

DEPARTMENT OF ARCHITECTURE AND CIVIL ENGINEERING
DIVISION OF APPLIED ACOUSTICS
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden 2025
www.chalmers.se

MASTER'S THESIS 2025

Acoustic Source Localization for an Indoor Pass-By Measurement System

A Beamforming Approach Using a One-Dimensional Sparse Microphone Array

Gustav Seger
Semir Sehic

CHALMERS
UNIVERSITY OF TECHNOLOGY

Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
Gothenburg, Sweden 2025

Acoustic Source Localization for an Indoor Pass-By Measurement System
A Beamforming Approach Using a One-Dimensional Sparse Microphone Array
GUSTAV SEGER
SEMIR SEHIC

Supervisors:	Frédéric Wullens	Volvo Group Trucks Technology
	Tor Möller	Volvo Group Trucks Technology
Examiner:	Jens Ahrens	Division of Applied Acoustics

Master's Thesis 2025
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology
SE-412 96 Gothenburg
Sweden
Telephone +46 31 772 1000

Cover: Delay and Sum beamformer output showing the array response of a one-dimensional sparse microphone array to a single acoustic source.

Typeset in LaTeX
Gothenburg, Sweden 2025

Acoustic Source Localization for an Indoor Pass-By Measurement System
A Beamforming Approach Using a One-Dimensional Sparse Microphone Array
Gustav Seger
Semir Sehic
Department of Architecture and Civil Engineering
Division of Applied Acoustics
Chalmers University of Technology

Abstract

Accurate source localization of acoustic sources is critical for vehicle noise analysis. This thesis examines the possibility of using sparse microphone arrays, which are part of an indoor pass-by measurement rig, as acoustic cameras in a near-field scenario. Four different techniques were implemented, validated and comparatively analysed. Three beamforming techniques, Delay and Sum (DS), Minimum Variance Distortionless Response (MVDR) and Functional Beamforming (FBF). The fourth technique is a post-processing algorithm, A Deconvolution Approach for the Mapping of Acoustic Sources (DAMAS). Each technique was evaluated based on spatial resolution, robustness to noise, array element imperfections and performance on real measurement data.

The DS beamformer demonstrated robustness to array element sensitivity variations and placement errors, but was limited under low Signal-to-Noise Ratio (SNR) conditions due to its broad mainlobe and elevated side-lobe levels. MVDR and FBF, both implemented using Cross Spectral Matrices (CSM), offered improved directional accuracy and noise suppression capabilities. FBF introduced a tunable order factor which enhances control of side-lobe suppression and mainlobe width, making it particularly useful as a pre-processing step for DAMAS. The DAMAS algorithm was implemented as a post-beamforming processing tool, offering improved spatial resolution, although its performance was based heavily on the quality of the beamformer input.

Validation included both simulations and real measurements conducted inside a semi-anechoic pass-by test rig, using stationary vehicle signals. All techniques successfully localized tonal sources across a frequency range of $40 - 1600$ Hz. However, the spatial resolution remained constrained by array aperture size and wavelength. Ground reflections introduced interference patterns that reduced accuracy at certain frequencies, mimicking the acoustic behaviour expected in a road-like environment.

The thesis concludes that combining MVDR and FBF for initial localization and DAMAS for source separation provides a flexible approach for acoustic source localization. The importance of accurate propagation and time delay modelling, calibration and signal processing was highlighted as a key factor to achieving reliable beamforming results in vehicle acoustic pass-by measurements.

Keywords: Beamforming, Acoustic Source Localization, Microphone Array, Delay and Sum, MVDR, Functional Beamforming, DAMAS, Acoustic Camera, Pass-By, Volvo

Acknowledgements

We would like to express our sincere appreciation to our supervisors, Dr. Frédéric Wullens and Tor Möller at Volvo Group Trucks Technology, for their guidance and support throughout this master's thesis. We would also like to thank Torbjörn Ågren and the additional staff in the NVH department for their assistance, helpful discussions, and for providing a collaborative and supportive working environment.

Lastly, we would like to thank our supervisor and examiner, Prof. Jens Ahrens, for his constructive feedback, insightful suggestions, and expertise in acoustic signal processing, which greatly contributed to the quality and depth of this work.

Gustav

I would like to thank my partner, Paula, for her support during my time at the university and throughout this master's thesis, which has been invaluable. I am also grateful to my family for their encouragement and support. Finally, I would like to thank my friends outside of the university, as well as my classmates, for their collaboration and discussions.

Semir

I want to thank my family for always supporting me throughout my studies. Their support has meant a lot during this period. I would also like to thank my friends and colleagues, both inside and outside the university, for their encouragement and collaboration during the thesis work.

Gustav Seger and Semir Sehic, Gothenburg, August 2025

Usage of AI Tools

Artificial Intelligence tools has been used for grammar and spelling correction, formatting of LaTeX tables and MATLAB figures. Also for generating a background image of a truck for some figures. No AI tools were used to generate any original text or code in this thesis.

List of Acronyms

This thesis use of acronyms is presented below with individual acronyms and an accompanying explanation.

AOA	Angle of Arrival
CSM	Cross Spectral Matrix
DAMAS	A Deconvolution Approach for the Mapping of Acoustic Sources
dB	Decibel
dB SPL	Decibel Relative to Sound Pressure Level
DFT	Discrete Fourier Transform
DL	Diagonal Loading
DOA	Direction of Arrival
DS	Delay-and-Sum
DS-CSM	Delay-and-Sum algorithm utilising a Cross Spectral Matrix
DSP	Digital Signal Processing
ESM	Equivalent Source Model
EVD	Eigenvalue Decomposition
FBF	Functional Beamforming
FIR	Finite Impulse Response
FRF	Frequency Response Function
GSM	Gauss–Seidel Method
HPF	High-Pass Filter
IDFT	Inverse Discrete Fourier Transform
IR	Impulse Response
ISM	Image Source Model
MPDR	Minimum Power Distortionless Response
MVDR	Minimum Variance Distortionless Response
NVH	Noise Vibration Harshness
PSD	Power Spectral Density
RMS	Root-Mean-Square
SNR	Signal-to-Noise Ratio
SPL	Sound Pressure Level
SRP	Steered Response Power
SULA	Sparse Uniform Linear Array
TDOA	Time Direction of Arrival
ULA	Uniform Linear Array

Contents

List of Figures

List of Tables

1
Introduction

This thesis presents the development and implementation of acoustic camera software for an indoor pass-by measurement system, using already existing microphone arrays located in a semi-anechoic chamber. The thesis is a collaboration between the Division of Applied Acoustics at Chalmers University and Volvo Group Trucks Technology, which will be referred to as Volvo throughout this thesis. The semi-anechoic chamber is located at the Volvo Noise & Vibration Laboratory in Lundby, Gothenburg. The semi-anechoic chamber is built according to specifications required to perform pass-by measurements, a measurement method essential for the automotive vehicle industry assessing noise levels emanating from vehicles. All necessary hardware is already in place, but as it is not the main purpose of the microphone arrays to serve as acoustic cameras no dedicated software currently exists. This thesis aims to design and implement this software, integrate it into Volvo's existing measurement infrastructure. Creating a powerful analysis tool for Volvo Noise, Vibration & Harshness (NVH) engineers. The aim of this platform is to enable detailed acoustic source localization and characterization using the same signals recorded during standard pass-by measurements testing procedures.

1.1 Background

The automotive industry relies on a standardized measurement method known as pass-by measurements to validate that vehicles conforms with international legal standard and noise regulations. These measurements are performed outdoors on designated test tracks for certification and regulatory approval but they can also be performed indoors [1] with the added benefits of having a stationary vehicle in a controlled environment, enabling detailed NVH research and development work. The procedures and requirements for both outdoor and indoor pass-by measurements setup is regulated in the standard ISO 362-1:2022 and 362-3:2022 respectively.

In the outdoor setup, a vehicle is driven along a test track and passes through a measurement area where microphones are positioned to capture the pass-by noise levels. This setup is designed to mimic the noise contribution of a vehicle passing in a urban environment [2]. The indoor version of this test aims to replicate the acoustic conditions of the outdoor scenario by requiring one close vicinity reflection [3]. For the outdoor measurements situation this reflection originates from the road surface, while indoors the reflection originates from a hard floor surface. These are typical sound field characteristics of a semi-anechoic chamber, which often features a hard, reflective floor surface combined with absorptive walls. Rather than using a test track and allowing the vehicle to physically pass-by the microphones in

the measurement zone, the indoor setup places the vehicle stationary on a dynamometer and uses fixed microphone arrays to simulate the pass-by motion acoustically.

In the Volvo Noise & Vibration Laboratory semi-anechoic chamber, a dynamometer is centrally placed with uniform linear arrays (ULA) of microphones placed along the long sides of the chamber, see figure 1.1 for a schematic model of the setup. Volvo NVH test engineers have identified the potential utilizing these ULAs as acoustic camera systems. Processing the signals from individual microphones for source localization purposes, and assessing the timbre of a vehicle under test. Such a system could serve as a powerful tool enhancing the engineers analysis ability while doing pass-by measurements and create a better understanding of the acoustic behaviour of the vehicle being tested within the semi-anechoic environment.

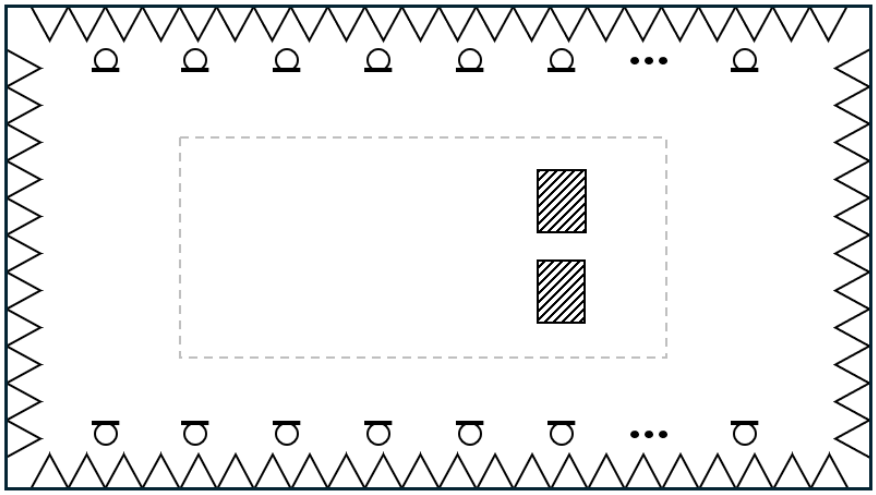


Figure 1.1: Schematic model of the semi-anechoic chamber located at the Volvo Noise & Vibration laboratory. Circles with extra line on top represents individual microphones. The two rectangles with solid hatching represents individual rollers of the dynamometer. The dashed grey rectangle represents the area in which a vehicle can be positioned.

1.2 Purpose and Aim

The foundations for this master's thesis work were established in a master thesis proposal provided by our supervisors at Volvo Group. They were then further refined in a subsequent meeting between Volvo Group supervisors and university supervisor from Chalmers present, at date $16/12 - 2024$. Based on the information from this meeting, a master's thesis application document and later a planning report was written. At date $8/4 - 2025$, a half-time meeting was held at the Volvo Noise & Vibration Laboratory between us and supervisors from Volvo Group and Chalmers. The information from these previous documents and meetings served as the basis for the final refinement of the purpose and aims for this master's thesis work, which are outlined in this section.

The overall objective of this master's thesis is to add additional purpose to the ULAs and the indoor pass-by measurement rig at Volvo Noise & Vibration Laboratory, enabling their use as acoustic cameras. The primary purpose is the development of necessary software, which

must integrate with the pre-existing measurement infrastructure currently use.

The existing ULAs in the semi-anechoic chamber can be considered to be sparse uniform linear arrays, abbreviated as SULAs, due to the wide spacing between individual microphones compared to the frequency range of interest. This limits their potential usefulness, utilising them as acoustic cameras. Exploring and using suitable signal processing techniques, the goal is to provide reliable source localization abilities within as much of this frequency range of interest as possible.

This can be translated and further condensed into research question that this master's thesis work aims to answer.

- What level of performance can be achieved by developing and implementing acoustic camera software for a microphone array that is a part of an indoor pass-by measurement rig?
- How do individual signal processing techniques perform in this context, and what improvements can be realized by combining them into a unified acoustic camera software?
- What is the viable frequency range for acoustic source localization achievable by the acoustic camera software, and how is this limited due to the sparsity of the microphone arrays?

1.3 Limitations and Demarcations

The ULAs are situated in a very controlled semi-anechoic laboratory environment and the finished acoustic camera should integrate as seamlessly as possible into the already existing systems in use. This puts some constraints on the system and the environment that it will be evaluated within. It also affect the possibility to take the findings from our work and applying them in a different situation. So the results from this work will be based on simulations and measurements, validated for use in a semi-anechoic environment with a ULA setup that is similar to the demands stated in ISO 362-3 for indoor pass-by measurements.

Equipment available during a standard pass-by measurement in the semi-anechoic chamber is what to be used and no altercations to any systems or setup should be needed in order to operate the acoustic camera software. This will make sure that the software is easily integrable into the measurement systems and won't require any additional setup for the NVH test engineers. It also has the advantage that the software can be used on previously gathered data and isn't restricted to measurement data gathered post-implementation.

The acoustic camera software will not perform real-time analysis or processing. Thus, the software has access to the whole signal content, it has knowledge of the content in the next sample and it knows the total length of the input signals. Due to how data is extracted from the current measurement system it is not feasible to integrate an external software into this ecosystem in a real-time capacity, as a part of this master's thesis. This ensures that our work and coding efforts remain relevant and is primarily focused on acoustic and signal processing related theory and technology, appropriate for this master's thesis work.

There are some acoustical quantities in terms of frequency range and angular coverage that are thoroughly defined which will remain throughout this master's thesis, they are summarized in the bulleted list below. The frequency range of interest was defined by our supervisors and Volvo NVH engineers, it represents a compromise between frequency range of interest for source localization purposes and feasibility due to technical and geometrical limitations. The angular coverage range is based on a possible source area, see figure 1.1, there is an area in which vehicles could be placed and this imposes limits on the valid frequency range.

The signals evaluated are based on input from Volvo NVH engineers and are primarily concerned with distinct tonal components in a vehicle-like timbre. ULAs only has microphone distribution in one dimension, therefore sources in other directions than the same plane of the ULAs has been disregarded. One measurement was conducted with source located at elevated height and is presented in subsection 8.2 but this has not been thoroughly investigated and remains as a limitation.

- Frequency range of interest is 50 Hz to 2 kHz.
- Possible source area angular coverage, based on the array's central global bearing angle, is approximately $-36°$ to $25°$, at 7.5 m distance from the array.
- The primary interest is achieving reliable source localization capabilities within these previously defined ranges.
- Signal types considered are mainly singular tonal components in a vehicle-like timbre.
- The primary focus is on sources located in the same plane as the microphone array's.

1.4 Thesis Structure

The overall structure and layout of this master's thesis work is centred around four techniques chapters. They all contain a description, a technique-specific theoretical background, implementation, validation metrics results and an analysis. The thesis chapter order from the start is a introductory chapter, followed by a fundamental theory and Concepts chapter which contain non technique-specific theory, methodology, technique-specific chapters, comparative results, discussion and conclusion. Further below is a bulleted list containing the chapters and a description of their content.

Four different techniques has been examined, individual algorithms has been developed and designed as part of the acoustic camera software. Where the synergy between them, and the combination of them, outlines the function and performance of the actual finalized software. The individual techniques consists of three beamformers. Delay and Sum (DS), Minimum Variance Distortionless Response (MVDR) and Functional Beamforming (FBF). The fourth technique is a post-processing algorithm, A Deconvolution Approach for the Mapping of Acoustic Sources (DAMAS). These techniques are presented one per technique-related chapter. These chapters contains a short description, a theoretical background explaining the theory related to that specific technique, a brief description in the design and implementation and how they fit into our ecosystem. The next section are validation metrics which aims

to quantify and measure each beamformers performance through simulations, this acts as a basis for the analysis which is an evaluation of each technique and a basis for choosing the next technique examine to implement.

- **Introduction**. Provides background, aims and purposes and relevant limitations.
- **Fundamental Theory and Concepts**. Introduces general concepts and theory needed for the understanding of the subsequent material.
- **Methodology** Describes the research work approach and the workflow applied throughout this process. Provides insight into the design of the software, simulations and measurements.
- **Technique chapters**. Four separate specific technique chapters, one for each implemented technique DS, MVDR, FBF and DAMAS. Each chapter includes a technique specific description, theoretical background, validation metrics used for performance evaluation and an analysis based on these validation metrics.
- **Comparative results**. A comparative presentation of results relating to each individual techniques and applies them to measurement data.
- **Discussion**. Interpretation of gathered results, both highlighting individual technique findings and the synergy between them.
- **Conclusion**. Summarizes and presents the main findings, and recaps the result in a condensed version. Gives suggestion for future work.

2
Fundamental Theory and Concepts

This chapter presents fundamental theory and important concepts that together forms a foundation of knowledge for understanding the techniques and methods presented in the later chapters of this master's thesis. More advanced technique-specific theory is introduced within each respective technique chapter.

2.1 Beamforming

Beamforming is a type of signal processing technique which can be applied to recorded signals of sensor arrays, like microphone arrays, to filter signals based on their direction of arrival (DOA). A sensor array consists of multiple sensors arranged in a known geometry which introduces spatial diversity. This enables beamforming algorithms to process the incoming signals in such a way that they enhance components from a desired direction while also attenuating signals from other directions without physically moving the array.

Today beamforming is widely used in many modern technologies. In wireless communication, medical ultrasound, radar, sonar and also consumer audio devices such as speakerphones and hearing aids. Beamforming is not limited to receiving arrays like microphones. The same principles apply to transmitting arrays consisting of loudspeakers. In this specific case, beamforming can be used to project sound pressure in specific directions.

The principle of beamforming relies on the time or phase differences that a wavefront experiences as it reaches each sensor of the array. By applying calculated time delays at each sensor so that signals arriving from a particular direction align in time and phase, the signals can be constructively combined while signals from other directions interfere destructively. This creates an amplified gain response in the target direction, forming a directional sensitivity pattern known as the array response or also referred as Steered Response Power (SRP). This makes beamforming a essential tool for applications such as acoustic imaging, speech enhancement and source separation.

2.2 Acoustic Camera

An acoustic camera is a measurement system used to visualize and localize sound sources in spatial domain. It combines an array of microphones with beamforming or other signal processing techniques to estimate the DOA of acoustical sources [4]. It combines DOA estimations to generate images that represent the spatial distribution of acoustic energy. This acoustic image, depicting the sound field emitted from acoustical sources, can then be superimposed onto a camera picture of the same source. The composite image represents the output from an acoustical camera measurement system.

It exists a plethora of acoustic camera designs. Some are stationary some are handheld, different microphone array configurations all adapted to fit a certain purpose. Figure 2.1 depicts company Norsonics acoustic camera Nor848B and the Nor848B user interface during a building acoustic related measurement of noise through an office glass-door. In the figure, it is possible to see that the noise emanates primarily from the door gaps.

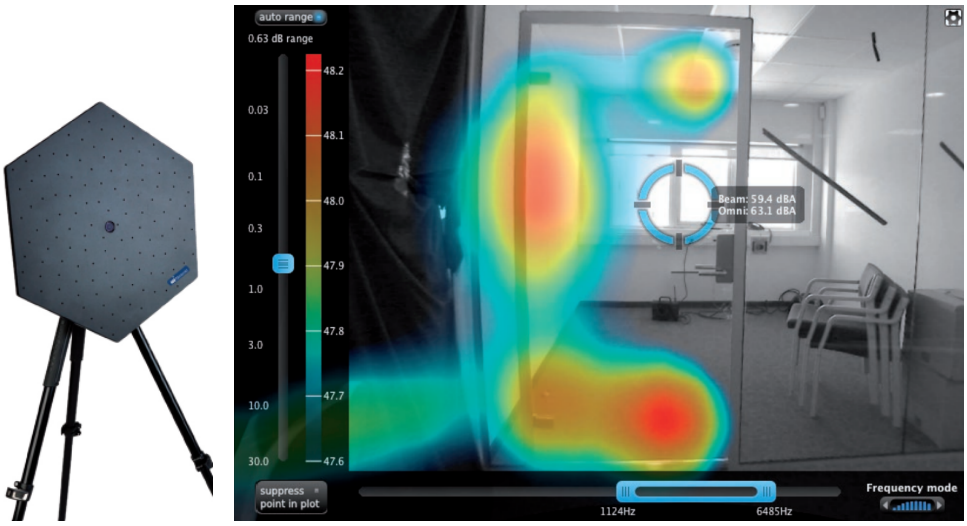


Figure 2.1: Acoustic camera Nor848B (left) and the user interface for the Nor848A system (right), as shown in a case study by Norsonic [5]. Images reproduced with permission from Norsonic.

2.3 Aliasing

The concepts of aliasing is central in all digital signal processing (DSP) applications and it is a phenomena which is relevant every time a signal is converted from analogue to digital domain, this conversion is usually expressed as sampling. In digital systems, the sampling rate f_s expresses how many times per second a system captures an incoming analogue signal and takes a snapshot of the input value at discrete time locations, this is known as time domain sampling. How often these snapshots are taken compared to the content of the incoming signal has large implications for the behaviour of the signal in the digital domain.

In order to avoid aliasing the *Nyquist sampling theorem*, or sometimes mentioned as *Nyquist-*

Shannon sampling theorem, needs to be satisfied. This theorem states that the sampling frequency f_s must be at least twice the maximum frequency component f_{max} in the incoming analogue signal in order to faithfully recreate the signal in digital domain [6] [7]. This can be expressed as seen in the equation below.

$$f_s \geq 2f_{max} \tag{2.1}$$

Spatial Aliasing

In addition to aliasing phenomena in time domain sampling, it is also a crucial factor for a microphone arrays valid frequency range. In this application it is referred to as spatial aliasing, which relates the geometry of the microphone array to a wavelength λ of a certain frequency and the previously described Nyquist theorem.

For a microphone array to properly represents an impinging wave there needs to be at least two microphones per λ [8]. This places a requirement on the geometrical distance d between individual microphones of the array relative to the highest frequency which it aims to capture. If this requirement is not satisfied, ambiguities in the array response arises in the form of grating lobes which can be seen as additional lobes visible on both sides of the mainlobe. Figure 2.2 shows a representation of a DS beamformer with side-lobes and grating lobes visible.

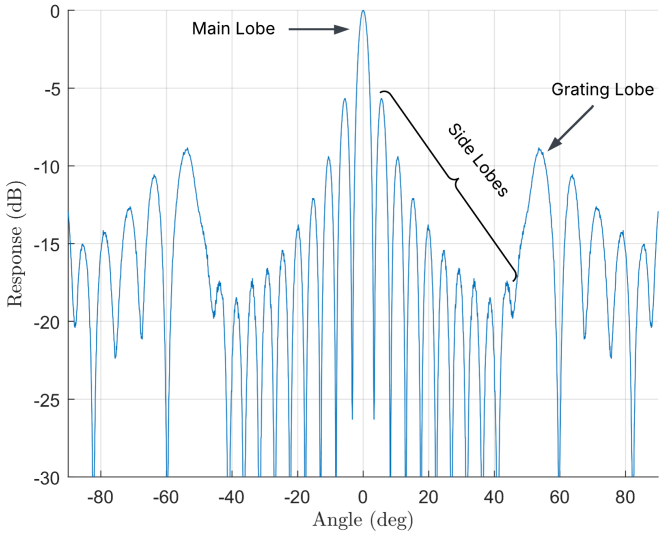


Figure 2.2: Array response showing mainlobe, side-lobes, and grating lobe.

The requirement of 2 microphones per λ is relevant in situations where the angle of arrival (AOA) can be anywhere in a semicircle of $-90°$ to $90°$. In setups where this is not expected, a relaxation of the requirement can be applied which ensures that the AOA range of interest is always within the first grating lobe appearances [9]. See equation 2.2 below where d is the equidistant spacing between microphones and θ should reflect the largest expected AOA in certain setup. Solving this equation for the frequency f (part of λ) provides a limit,

below this frequency no grating lobes appear. The frequency denotes the first introduction of grating lobes into the array pattern for a certain microphone array setup.

$$d \leq \frac{\lambda}{1 + |\sin\theta|} \tag{2.2}$$

Equation 2.2 provides a geometric condition for avoiding spatial aliasing in a sensor array. Ensuring that grating lobes do not enter the visible angular range. By rearranging equation 2.2 we can express this relationship in terms of frequency:

$$f = \frac{c}{(1 + |\sin\theta|)d} \tag{2.3}$$

In addition to estimating frequency at which grating lobes appear we can also approximate its angular position using the following relation derived from the geometry and phase difference between elements [9]:

$$\theta = \sin^{-1}\left(\frac{m \cdot 2\pi + \Delta\phi}{2\pi} \cdot \frac{\lambda}{d}\right) \tag{2.4}$$

where m is the grating lobe index (e.g, $\pm 1, \pm 2, ...$), $\Delta\phi$ is the phase shift per sensor element, λ is the wavelength and d is the distance between sensors in the array.

The number of microphones in an array directly affects its spatial resolution. With more microphones resulting in narrower mainlobe and improved angular resolution. Figure 2.3 shows the array response of DS beamformer configured with 8, 16, 32 and 64 microphones. As the element count increases, the mainlobe width narrows, which enhances the ability to distinguish between sources with small angular separation.

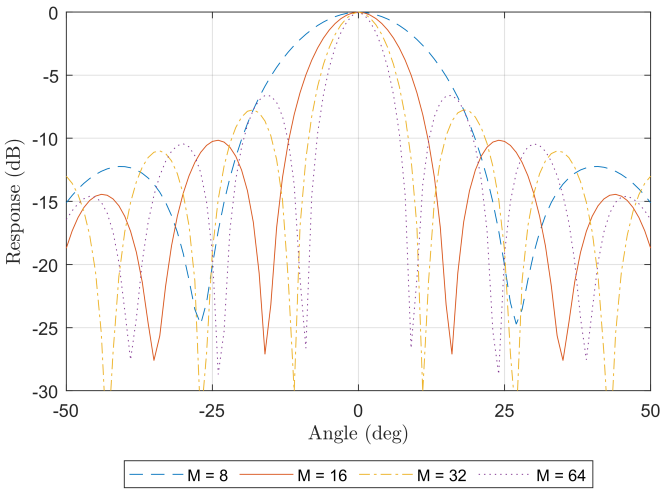


Figure 2.3: Array responses of DS beamformers with 8, 16, 32, and 64 microphones, illustrating the effect of element count on angular resolution. The source is a 100 Hz tone positioned on-axis relative to the arrays.

2.4 Microphone Arrays

A standard pressure type microphone senses changes in air pressure and converts the mechanical movement of its diaphragm into an electrical signal, which is transferred to auxiliary equipment in a signal chain to interpret.

Using a single standard pressure type microphone to capture a specific acoustic sound field does not provide sufficient information needed to estimate either DOA or the distance between source and microphone. For example, analysing just the dB SPL of a captured signal, a conversation in close proximity at 65 dB SPL would register similar levels as a concert at 100 dB SPL at 60 m distance! Thus the single microphone signal does not contain valid information for spatial source localization in itself. Instead using multiple microphones arranged in an array introduces timing differences, as a specific wavefront reaches each individual microphone with a slight delay, see figure 2.4 for an illustration. These timing difference are often called time difference of arrival (TDOA) and they can be exploited to estimate DOA of a certain source.

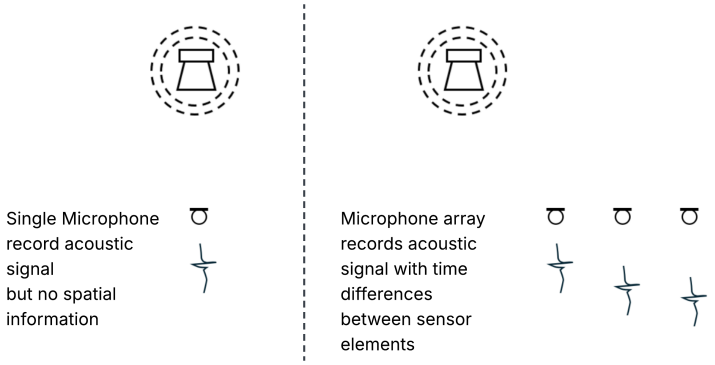


Figure 2.4: Visualizing TDOA for a single microphone and a array with arbitrary impulse like source signal.

A microphone array consists of multiple microphones used in conjunction with each other to capture information about acoustical sources, sound fields and their environment [6]. By analysing the differences in the signals received at each microphone, such as time delays, phase shifts and amplitude differences it becomes possible to estimate the DOA of sound waves and spatially localize sources with the help of signal processing techniques like beamforming. These arrays can be configured in various geometries, each optimized for different types of measurements and environments. The choice of array configuration depends on several factors including the desired spatial resolution, frequency range of interest and understanding whether the sources are located in the far or the near-field of the array.

Uniform Linear Array

Uniform Linear Arrays (ULAs) are one dimensional arrays consisting of multiple elements placed with equidistant spacing along the the same axis [10]. The type of elements in the array depends on the application, for this acoustic camera implementation the focus is on hearable sound and then the individual elements should be microphones. Figure 2.4 illustrates

a 3 microphones ULA setup on the right hand side with equidistant spacing between the microphones.

Sparse Uniform Linear Array

Sparse Uniform Linear Arrays (SULAs) are ULAs impacted by spatial aliasing due to the distance between individual elements. See subsection 2.3 for a explanation of the spatial aliasing concept.

The sparsity of a certain ULA is defined by the number of elements that are placed within half a wavelength $\lambda/2$ spacing of each other compared to the frequency of interest. The definition used by Patwari in *Sparse Linear Antenna Arrays: A Review* of a sparse array is grid-based but it can be translated simply to a frequency based interpretation. In the grid based version every grid-point equals $\lambda/2$ distance, which is the demand for no spatial aliasing to occur at a certain frequency. If there are fewer elements then the number of grid-points, with first and last grid-points still occupied, the array can be considered to be sparse [10], or a SULA in this case. The frequency based interpretation of this is that the distance between elements in the array needs to be shorter than the λ of the frequency of interest so that there are at least two microphones per λ, else spatial aliasing occurs and grating lobes appears.

2.5 Near- and Far-Field

Differentiating between near-field or far-field setups are critically important for beamformers that relies on a steering vectors in order to steer the response towards a certain angle or location. If the correct assumption is not made it will hinder the algorithms ability to beamform and produce a valid result, the beamformer won't be able to aim.

Consider an omnidirectional sound source emitting a signal. If a microphone array is positioned close to the source, the SPL at each individual microphone would vary significantly, the SPL distribution over the microphone array would be uneven. This occurs due to the spherical characteristics of the wavefront impinging on the microphone array. If the microphone array was instead placed at a greater distance from the source, the SPL distribution over the microphones would be similar to each other. At larger distances, the spherical wavefront has expanded so much that the impinging wave on the microphones has planar characteristics which results in similar SPL distribution over all microphones.

The distinction between near-field and far-field is primarily based on the spatial characteristics of the wavefronts emitted by a source. Far-field conditions is assumed when wavefronts impinging on a microphone array are approximately planar, which happens where distances from the source to the microphones array are sufficiently large. At smaller distances the wavefront retains its spherical characteristics and the microphones SPL varies by a larger margin.

Determining whether a specific setup operates within near- or far-field conditions can be approached in numerous ways. There are theoretical criteria but the evaluation can also be done practically based on geometric intuition or by recording the signal of a source and

evaluating the response of the microphone array.

A theoretical criterion for the transition between near- and far-field regions can be found in antenna literature, where its defined for transmitting antennas. The boundary is often referred to as the Rayleigh distance, or alternatively as the Fraunhofer distance [11]. The expression describes at what distance d from a transmitting antenna, with array dimension D and at a certain frequency with wavelength λ, the transition between near- and far-field happens [12], see equation 2.5. Applying this expression on a microphone array, d describes at what length from the array the near- far-field transition happens for a certain λ. To assess if a microphone array is within near- or far-field d has to be compared to a source position.

$$d = \frac{2D^2}{\lambda} \tag{2.5}$$

2.6 Wavefront Modelling for DOA Estimation

Accurate DOA estimation relies on the assumptions made about the geometry of the incident wavefronts. Beamforming algorithms vary depending on whether the microphone array is located in the near-field or the far-field. Conventional beamforming techniques typically assume far-field conditions, under which the wavefronts arriving at the array is regarded as planar [13]. This implies that the DOA is approximately constant across all microphones.

Beamforming under far-field assumptions becomes inaccurate when the microphone array is placed close to the source. In near-field the wavefronts exhibit noticeable curvature and the planar approximation no longer holds. Instead the sound pressure field at the array should be considered as spherical wavefronts, arriving at each microphone with different delays and amplitudes [14]. Understanding the differences between planar and spherical wavefronts is essential for selecting the appropriate propagation model in beamforming and is a crucial part of our beamforming algorithms and simulations which should mimic the behaviour of sound in the semi-anechoic chamber.

Plane Wave Model

In far-field conditions, where the sound source is located at a large distance relative to the array aperture and acoustic wavelength. Incident wavefronts can be approximated as planar, this assumption also simplifies acoustic modelling. A plane wave propagating in the $+x$ direction can be described by the following expression for the sound pressure at position x and time t:

$$p(x,t) = Ae^{-jkr}e^{j\omega t} \tag{2.6}$$

where A is the signal amplitude, k is the wavenumber, ω is the angular frequency, and $r = x$ for one-dimensional propagation.

In this model the wavefront arrives at the array with a uniform angle of incidence, resulting in approximately equal DOA across all microphones. The amplitude is assumed to remain constant across the array and the phase difference between microphones depends only on the

relative positions and the angle of arrival [13]. Figure 2.5 illustrates the geometry of a ULA with an impinging planar wavefront.

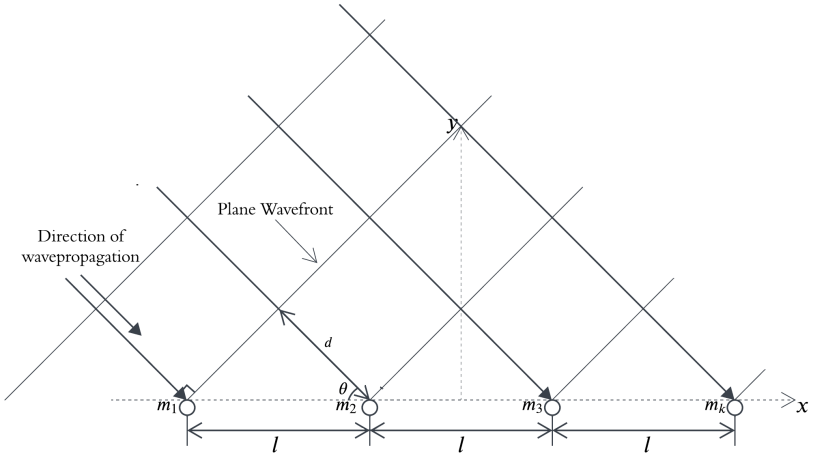


Figure 2.5: Propagation of plane wave and line array

Assuming a plane wavefront and far-field conditions, the additional path length to the mth microphone (relative to the first microphone) is given by:

$$d = (m_k - 1) \cdot l \cdot \cos(\theta) \tag{2.7}$$

where d is the relative path difference to the mth microphone, l is the inter-element spacing and θ is the DOA. From Equation 2.7 the corresponding time delay τ_p at the mth microphone is:

$$\tau_p = \frac{d}{c} \tag{2.8}$$

where c is the speed of sound in air.

Spherical Wave Model

In the near-field, the curvature of the wavefront must be accounted for when modelling acoustic propagation. A point source located at a finite distance from the array produces spherical wavefronts, meaning that both the amplitude and phase of the received signal vary across microphones depending on their relative distances to the source. The sound pressure signal at the mth microphone, $p(r,t)$, is modelled as a spherical wave propagating from a point source. Assuming only the direct sound path and neglecting reflections (e.g., from the ground or other surfaces), the pressure can be expressed as:

$$p(r,t) = \frac{A}{r_{m_k}} e^{-jkr_{m_k}} e^{j\omega t}, \tag{2.9}$$

where A is the complex amplitude of the wave, r_{m_k} is the distance from the source to the mth microphone, k is the wavenumber, and ω is the angular frequency. The term $e^{-jkr_{m_k}}$

accounts for the phase delay due to spherical propagation.

It is important to highlight the significant difference in measured signals depending on the type of wavefront, spherical (Eq 2.9) or planar (Eq 2.6). If an algorithm assumes planar wavefronts but the actual incoming waves are spherical. This mismatch introduces errors in the signal processing and estimation. Proper modelling of spherical wavefronts is essential to ensure precise results if an microphone array is located within the near-field region.

The sound pressure signals received at each microphone exhibit both amplitude and phase differences due to the propagation characteristics of the wavefront. For a monopole source the amplitude factor at the mth microphone is given by:

$$\frac{A_+}{r_{m_k}}$$

where r_{mk} represents both the physical distance between source and microphone but also the distance attenuation of the amplitude A. While the phase difference between individual microphones is visible in the exponents of the spherical wave formula and can be expressed as:

$$-\frac{r_{m_k}}{c} + t$$

Figure 2.6 show the geometry of spherical wavefronts impinging on a ULA, which introduce Varying DOAs due to the curvature of the wavefront. The bearing angle θ_c is scanned within an angular range of 0° to 180° or 90° to -90°, depending on the chosen reference frame.

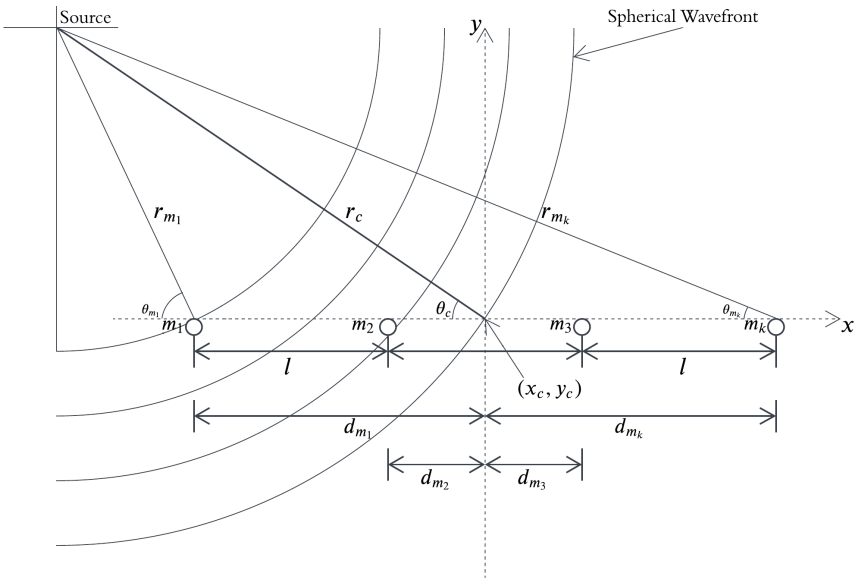


Figure 2.6: Propagation of spherical wave and line array.

The bearing angle θ_c can be calculated using the law of cosines which accounts for the distances between the array center, the microphones and the focus point

$$\theta_c = \cos^{-1}\left(\frac{r_c^2 + d_{m_k}^2 - r_{m_k}^2}{2 \cdot r_c \cdot d_{m_k}}\right), \tag{2.10a}$$

where r_c is the distance between the array center and the source (or focus point), d_{m_k} is the distance between the mth microphone and the array center and r_{m_k} is the distance between the mth microphone and the source (or focus point). If we shift r_c and r_{mk} from equation 2.10a we then compute the angle of each individual microphone by:

$$\theta_{mk} = \cos^{-1}\left(\frac{r_{m_k}^2 + d_{m_k}^2 - r_c^2}{2 \cdot r_{m_k} \cdot d_{m_k}}\right). \tag{2.10b}$$

Once the bearing angle of the source relative to the microphone array is determined it can be used to compute the relative propagation distances for each microphone. These distances directly influence the time delay at each sensor.

2.7 Image Source Model

The Image Source Model (ISM) is a room acoustics method used to calculate and estimate contributions from reflections in a sound field by treating each reflection as individual sources. With knowledge about room geometries and surface materials image sources can be virtually placed in the vicinity of room boundaries in relationship to an original sound source position and its signal. Early implementations, like Allen and Berkley, uses ISM in order to simulate impulse responses (IRs) of small rooms [15].

In this setup ISM is not used to emulate IRs but instead for simulation purposes. The environment where our microphone arrays are placed is semi-anechoic. The walls and roof are non-reflective. The floor is reflective and its reflection is expected to contribute to the overall sound field. Due to this very controlled environment there will only be one reflection present the semi-anechoic chamber. This makes ISM a potent candidate for simulation purposes since all signals will propagate along two paths, the direct path and the floor-reflected path, and can be summed at a microphone location as:

$$p_{\text{tot}} = p_{\text{dir}} + p_{\text{refl}} \tag{2.11}$$

Both p_{dir} and p_{refl} is pressure contributions from direct and reflected signal paths between a source and a microphone position. Equation 2.9 represents a complex spherical wave spreading from a monopole and updates the pressure field equation 2.11 at the microphone location

$$\begin{aligned} p_{\text{tot}} &= \frac{A}{R_{\text{dir}}} e^{-jkR_{\text{dir}}} e^{j\omega t} + Q\frac{A}{R_{\text{refl}}} e^{-jkR_{\text{refl}}} e^{j\omega t} = \\ &= A\left(\frac{1}{R_{\text{dir}}} e^{-jkR_{\text{dir}}} + Q\frac{1}{R_{\text{refl}}} e^{-jkR_{\text{refl}}}\right) e^{j\omega t} \end{aligned} \tag{2.12}$$

The distances R_{dir} and R_{refl} are visualized in figure 2.7 and how the gray image source is positioned outside of the room boundary in order to achieve the correct time delay related to the longer reflected distance. Q is the floor reflection factor. Equation 2.12 is complex, a real world recorded microphone signal captures the real part, therefore $p_{\text{meas}} = \Re(p_{\text{tot}})$. This simulation technique can be used for both noise and tonal sources.

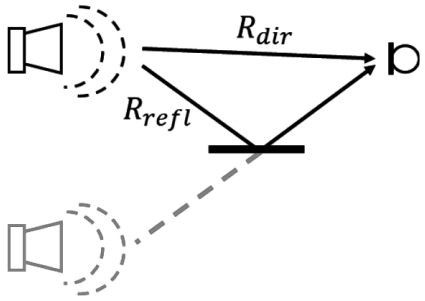


Figure 2.7: Source and image source contribution at a microphone position.

2.8 Time, Frequency and Spatial Domains

A domain in signal processing represents certain ways of visualizing the content of a specific signal. The standard domains are the time domain and frequency domain. The time domain captures how a signal changes over a period of time, while the frequency domain captures the frequency content of the signal. A specific sampled signal can be represented in both domains and be transformed between them using the Discrete Fourier Transform (DFT) and its inverse (IDFT).

Figure 2.8 illustrates a sinusoidal source in time and frequency domain representation. Between these domains, the grey box represents a transformation stage, showing how the DFT converts a time domain signal into its frequency domain counterpart. In the opposite direction, the IDFT is used to recover the time domain signal from the frequency domain representation. The DFT and IDFT can be expressed as in equations 2.13 and 2.14 respectively. Here, $x(n)$ represents a time domain signal expressed in samples n, $X(k)$ is the frequency domain representation of the same signal, where k is frequency bins. N is total number of samples in the signal.

$$X(k) = \sum_{n=0}^{N-1} x(n) e^{-j\frac{2\pi}{N}kn} \tag{2.13}$$

$$x(n) = \frac{1}{N} \sum_{k=0}^{N-1} X(k) e^{j\frac{2\pi}{N}kn} \tag{2.14}$$

In figure 2.8 there is an additional spatial domain included, which is important in beamforming applications. The spatial domain represents a geometrical relationship between a source and a microphone array, this introduces timing differences which the individual microphones captures. For acoustic cameras, the spatial domain describes how a recorded signal can be associated with a specific location in the analysed environment. Spatial domain representations of signals can be seen in figure 2.1 where a estimated source location is placed within an environment. Figure 2.2 pictures an array response which is another type of visualization of the spatial domain, here a source location is represented in angles relative to the microphone arrays global bearing angle.

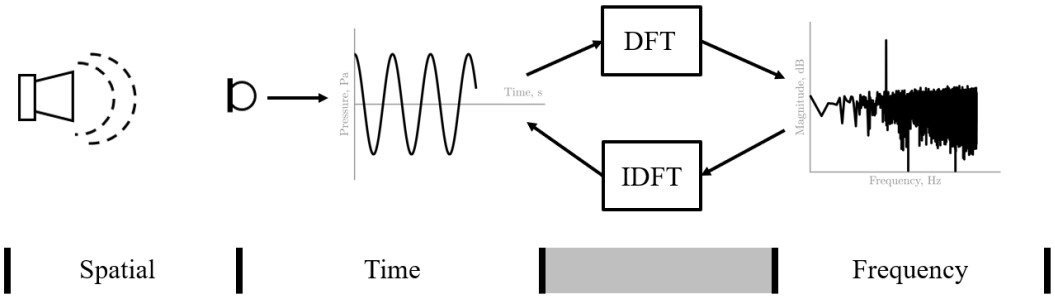


Figure 2.8: Representation of domains.

3
Methodology

This chapter presents the methodology and design process applied in this thesis and for the development of the acoustic camera software. It describes how techniques were selected based on intermediate results. It introduces the simulation design for both spatial domain environmental scripts and signal generation scripts. Also measurement design, used to validate technique performance and behaviour. Finally, the Validation metrics design for simulated verification of implemented techniques using Monte Carlo simulations.

The design process applied for implementing, validating and selecting the next technique was iterative and cyclic. Each iteration of the process started with establishing a theoretical foundation, followed by simulations. Measurements were performed when deemed necessary to gain further insight and to validate the simulated results. The final stage of each iteration was a quantitative analysis to assess the performance of a specific technique using simulated data and validation metrics. These were based on Monte Carlo simulations where a specific parameter was randomized across many trials in order to a achieve a robust and reliable performance metric result. The concluding analysis interpreted the validation metric results together with additional data and information gathered, and served as the basis for selection of the next technique to implement, with focus on achieving synergy between, and to improve the overall performance of the acoustic camera software. This approach was appropriate for the purpose of this thesis work. It ensured flexibility within the design process based on intermediate results and allowed for a combination of suitable techniques, based on their individual inherent strengths, weaknesses, performance and reliability.

All coding and development work related to this master thesis has been performed in MATLAB. The final acoustic camera software is a MATLAB environment based program and contains a number of scripts which can be used in conjunction, or chosen specifically based on what the user wants to evaluate.

3.1 Workflow

This subsection offers an in-depth description of the iterative and cyclic work process. This puts emphasis on the individual steps within the iterative process. To content varied per iteration but the main purpose always stayed the same. The process is visualized in figure 3.1 where the theory, simulation and analysis/conclusion were part of every iteration. Steps placed in parenthesis are not directly part of the process even though they are strongly related. The measurement step was performed based on availability of the semi-anechoic chamber and if additional validation was deemed necessary. The defined goals and limitations was expected to be rigid and inflexible and not change drastically during the course of the master's thesis. This has proven to be true and no continuous changes has been done, these parts have remained essentially the same with tweaks done for clarification purposes. The grey dashed arrows represents which previous steps in the iterative process is affected by decisions taken in the analysis/conclusion stage, it also highlights how the process is restarted.

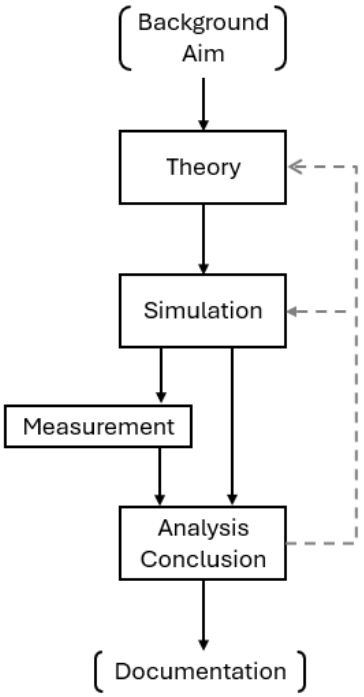


Figure 3.1: Flowchart visualizing iterative cyclic process.

Four techniques have been examined and evaluated as part of this master's thesis work, so the work process has been repeated as many times. The technique-specific chapters are 4, 5, 6 7. The techniques which has been evaluated are three beamforming techniques DS, MVDR, FBF. Also a post-processing algorithm, DAMAS. Below is a detailed description of the general content of each of the iterative process steps.

Theory

Examining the theoretical background related to a specific technique, deciding if it is viable in this specific context and if so, using the knowledge gathered here as a foundation for the upcoming steps in the process. This step directly affects how simulations are designed and for later stages evaluation.

Simulation

Development and implementation of simulation environments for a specific technique. The first simulations was always highly idealized and used to verify that techniques exhibited the expected performance in controlled scenarios. As insights were gathered, additional complexity was gradually introduced in order to better mimic realistic conditions. Typical additions were addition of external noise to emulate different Signal-To-Noise Ratios (SNR), background noise from the semi-anechoic chamber, varying source positions close at the extreme angles of the possible source area.

The information gathered in this stage regarding individual techniques performances was then later used in the analysis/conclusion-stage for performance evaluation and as a basis for deciding which techniques are viable for further implementation. This step was also crucial to decide if measurements was deemed necessary in a certain scenario and to decide the content of them.

Measurements

This process step was highly dependent on the preliminary results and findings from the simulation process step but also the availability of the semi-anechoic chamber and the possibility to perform measurements. Due to this, measurements were performed on a need-to-know basis. If the situation required measurements to validate results or simulation behaviour and also to find information not accessible through simulations. Data gathered from measurements accumulated and could be reused, reducing the need for further measurements at later dates as the thesis progressed. At the same time Volvo had a vast measurement archive which could be utilized.

The measurements that were performed during the period was heavily based on information retrieved during previous steps in this process and the measurement design always sought to validate or invalidate previous gathered simulated results or test hypothesis which helped us to increase our understanding.

At the start of the master's thesis work period, the intention was to perform one or two extensive measurement sessions, thoroughly planned with clearly defined goals and detailed measurement schedules. Due to the scheduling of the semi-anechoic chamber it was not possible to secure time slots with sufficient lead time for this approach to be viable. Instead, time slots tended to become available on shorter notices. This affected how measurements could be performed and in response, adapting to the scheduling, we chose to maintain a list of possible useful or needed measurements. The contents of this list was continuously updated. Some proposed measurements were at later dates deemed unnecessary, alternative ways of simulating or evaluating outcomes were developed, eliminating the necessity of some measurement. This approach ensured that only critical measurements were performed. At the same time, this approach had the unfortunate consequence that the limited lead time for us to prepare affected the quality of some measurements in unfortunate ways, sometimes redeemable, sometimes not.

Analysis and Conclusion

The final step in this iterative process main purpose is evaluation and decision-making. Data from simulations, measurements and insights from the overall process is gathered and compared. The combination of the information aims to evaluate and assess a specific techniques performance, both its advantages and areas where sufficient performance was not achieved.

As the master's thesis work progressed, the information available for decision-making increased, and its quality improved through more advanced simulation environments, greater amount of gathered measurement data, and general knowledge gained over time. Data from

previous iterations of specific techniques was also available and considered. The analysis conducted here is based on the estimated performance of one specific technique but also the synergies compared to previously implemented ones. This forms the foundation for determining which techniques is viable to study and implement in the next iteration.

3.2 Simulation Design

Throughout the working process for this master's thesis, simulation has been crucial for development and evaluation of the implemented techniques. The complexity of the simulation environments progressively increased as more information was gathered and increased accuracy was needed for proper evaluation. The simulation framework consists of multiple MATLAB scripts. They are modular and different combinations of them can be used to simulate scenarios more idealized or realistic depending on the goal of the simulation.

For presentation purposes, the different simulation scripts have been divided into two categories. **Environmental**, which contains scripts related to the coordinate system of the room and spatial domain localization. The second category is **Signals** and consists of scripts which emulates signals impinging on the microphone array and combines both signal creation and their expected behaviour at microphone locations.

3.2.1 Environmental

This category of scripts handles spatial domain information about positions and coordinate system within the semi-anechoic chamber. The systems part of the measurement rig in the semi-anechoic chamber has an internal coordinate system which the acoustic camera software needs to interpret correctly. At the same time the beamformers relies on a ULA center position form where it relates global bearing angle which does not coincide with the internal coordinate system.

initULA

The name is short for "Initialize ULA". This function defines microphone positions and places them correctly in the Volvo coordinate system. The function takes as input the total number of microphones, a microphone spacing, related to the equidistant spacing of microphones in the ULAs, and an input position of the first microphone in X, Y and Z coordinates of the room. Additional presets "left" and "right" can be passed as input which relates to both long side of the semi-anechoic chamber ULAs.

The output from this function is a list of 3D microphone position coordinates in the Volvo Cartesian coordinate system. See figure 3.2 where output from `initULA` is seen as the microphone positions.

gridBuilder

This script generates positions for the beamformers to aim at. It generates 2D spatial pixel maps based on predefined Volvo Cartesian coordinate system and desired pixel resolution in x and y directions. The output is distances and corresponding scan angles for the different beamformers.

The input modes are "source_area" which is defined as the possible source area with a small buffer added on every side. The "center2array" mode is an area starting just above the zero line seen from the ULA at use, and then places pixels in a pattern approaching the ULA. The last mode is "slits" which is pixels in a row at one specific distance, the main purpose of this is to be used with DAMAS, which can be read about in subsection 7.2.1 and 7.3.1, but it is also a valuable, simplified and fast analysis tool usable if the source distance is available or if just a specific distance is of interest.

Figure 3.2 shows output from `gridBuilder` as blue dots which are the pixel centers. The resolution in the figures is 25 pixels in x-direction, and 5 pixels y-direction, except for the "slits" mode which only has 1 pixel in y-direction.

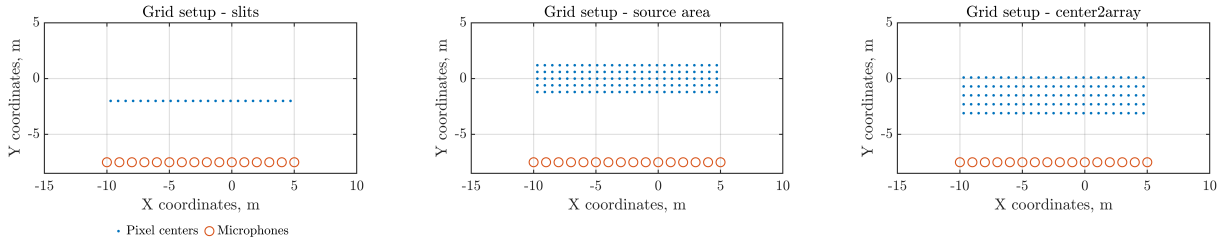


Figure 3.2: Output from environmental scripts, gridBuilder and initULA.

From the defined scanning area and the selected number of pixel positions in the x and y directions the script computes the corresponding spatial resolution between pixels. This is shown in figure 3.3 where the Cartesian grid layout is shown but here, corners of each pixels are highlighted, *dx* and *dy* represents the width and height of a specific pixel.

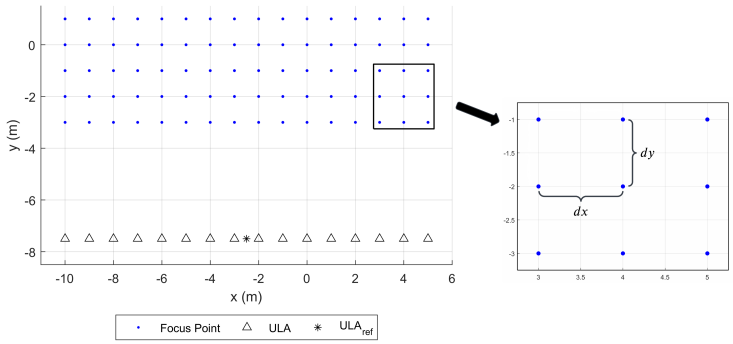


Figure 3.3: Pixel grid geometry resolution *dx* and *dy* for a defined pixel grid.

3.2.2 Signals

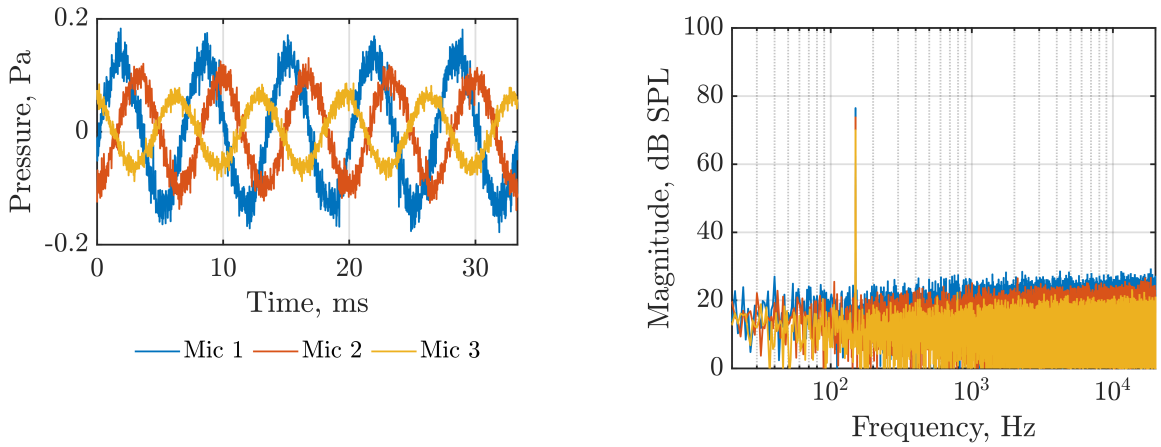
The scripts presented in this section are related to signal creation and their behaviour in relationship to the microphone arrays. It consists of three different scripts, a larger signal creation environment which creates sinusoidal and noise signals. Two simpler noise addition scripts which adds noise based on SNR and background noise from the semi-anechoic chamber.

simulateSphericalWave

This is the signal creation environment which has served as the foundations for all simulations. It has been refined continuously as demands on the simulations has changed. The initial basic script modelled sinusoidal tonal sources in a fully anechoic environment. At present the script can model both sinusoidal and broadband noise sources, in anechoic and semi-anechoic environments.

A key design challenge in the early development of the simulation environment was to understand if the microphone array was within the near- or far-field region in relationship to the possible source region, as this decides if a plane or spherical wave propagation model was suitable and as the name of the script suggests, a spherical propagation model has been implemented. Just by ocular investigation of the spatial setup and the geometrical relationship between the possible source area and the distribution of the ULA, it is reasonable to suspect that this is a near-field situation. Figure 3.2, central plot, illustrates the approximate extent of the possible source area but it is depicted with an extra size buffer, making the source area larger than any vehicle would be. The possible source area should be confined in x-direction between microphones 3 and 12, counted from left side of the plot. The spatial distance from any position within this area is small compared to the 16 m extent of the ULAs and it is therefore reasonable to draw the conclusion that a spherical propagation model is valid. Further examining the Rayleigh distance introduced in equation 2.5 also confirms that the possible source area is within near-field for all frequencies in the range where working in 50 Hz to 2 kHz. At the lowest frequency $f = 50$ Hz the distance between near- and far-field transition is $d = 75$ m and continues to increase with frequency. Thus spherical wave modelling should be applied since it correctly simulates both near- and far-field ranges.

The implication of this is that signals should be created as equation 2.9 suggest, a spherical propagation pattern. Sinusoidal signals is emulated exactly as the formula suggests with r_{m_k} calculated per microphone location and a provided source location. The real part of this complex signal is extracted as output, since it is the real part of the signal which relates to the sound pressure that a microphone records. White noise sources is simulated using the Matlab command `randn` with applied distance attenuation using r_{m_k} in the same fashion as for the sinusoidal signal. Figure 3.4 shows an example output of a simulated noise and sinusoidal source for an arbitrary 3 microphone ULA.



(a) Time domain signals.

(b) Magnitude spectrum.

Figure 3.4: Simulated sinusoidal and noise signal impinging on a 3 microphone array in an anechoic environment.

Semi-anechoic simulation behaviour is based on ISM as introduced in subsection 2.7 and it is enabled through an ISM mode input into the function. Analysing equation 2.12 the direct wave p_{dir} is already calculated in the fully anechoic simulation. Thus in order emulate semi-anechoic conditions the reflected wave calculation p_{refl} needs to be calculated and in order to do so the floor reflection factor Q and reflection path distance R_{refl} is needed.

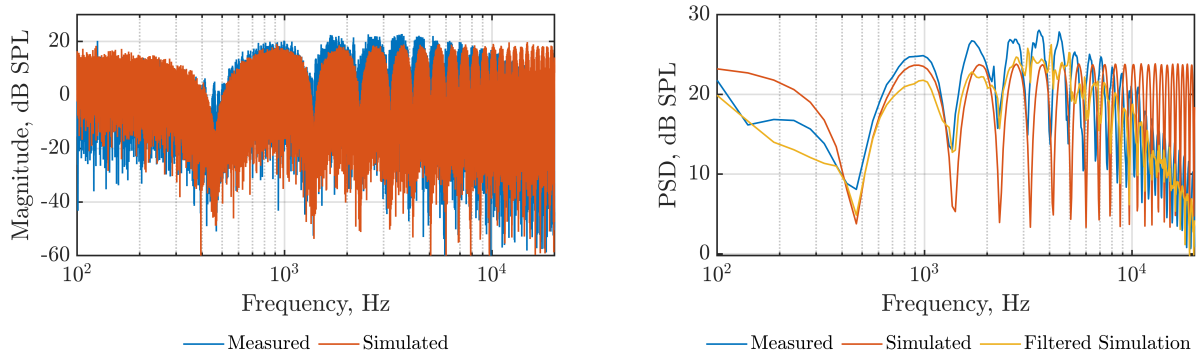
The ISO standard 10844:2021 specifies requirements for test track noise measurements, which indoor pass-by measurements rigs also must satisfy, in accordance with ISO 362-3:2022 [3]. The standard 10844:2021 states that an arithmetic average over 1/3-octave bands 315 Hz to 1.6 kHz should be calculated for numerous measurement positions and 80 % of these locations should have sound absorption coefficients α lower than 10 % [16]. It is deemed sufficient to set Q according to the mean of α as a reasonable approximation for all frequencies as

$$Q = \sqrt{1 - \overline{\alpha}} = \sqrt{1 - 0.05} \tag{3.1}$$

The expected reflection originates from the floor, therefore the image source position is set to the same 3D coordinates as the source but with its z-axis coordinate, height over floor, set to negative instead. This has the implication that $R_{\text{refl}} > R_{\text{dir}}$ thus achieving a small timing delay between p_{dir} and p_{refl} achieving constructive/destructive interference at different frequencies. This manifests itself as a comb pattern in a magnitude response. Figure 3.5 compares a measured and a simulated noise signal represented as a magnitude spectrum and using Power Spectral Density calculated via Welch's method using the Matlab `pwelch` command. The figures lower frequency limit is 100 Hz and does not fully cover the frequency range of interest, which extends down to 50 Hz. This is due to uncertainties regarding the frequency response of the loudspeaker and background noise levels, this representation ensures that the illustrated range is valid.

Experiments were conducted using additional filtering of the simulated signals to improve their agreement with measurements, a filtered simulation is illustrated as the yellow curve in figure 3.5b. Through simple amplitude modulation, the simulated signal levels could be fitted

to the same amplitude levels exhibited by the measured signals. This approach would ensure simulated amplitude levels similar to measured amplitude levels at locations where measurement data was available. Extending this amplitude modulation filtering to non-measured positions using interpolation techniques were proven difficult and results were often unstable. The filtered simulation visible in the PSD plot represents one of the most stable filter outcomes available. Applying the interpolated filtering on simulations at locations further from the measured positions (which acted as a basis for the interpolation) frequently exhibited erratic behaviour, no reliable compensation method was found or developed. Because of this, the original semi-anechoic simulation, with adapted Q and R-distances, were deemed to be sufficient. This simulation tends to overestimate levels compared to the measurements at lower frequencies than 200 Hz which is also obfuscated due to uncertain level of loudspeaker drop-off at lower frequencies. The simulation also tends to underestimate levels in the frequency range above 200 Hz and up until 2 kHz which is the upper limit of the frequency range of interest. Comparing dB SPL levels of the measured and simulated noise signals provide similar levels. At locations were measurements were available, there was no need to apply simulations. The noise signals presented here are primarily used to provide an analytical foundation for this subsection and not for validation or further simulation purposes.



(a) Magnitude Spectrum **(b)** PSD.

Figure 3.5: Magnitude spepctrum and PSD of measured and simulated noise signals. Additional filtering technique visible in PSD plot.

addNoise2Mic

This function adds white noise to input microphone signals and returns a noisy version, which emulates internal noise within the system. The inputs for the function is signals in a matrix format and a desired Signal-to-Noise Ratio (SNR) in dB. The function calculates the power of each individual signal and generates individual white noises per signal which is scaled to achieve the desired SNR value. The output is the combination of the input signals and the additional scaled white noise, which can be considered as additional added system noise.

backgroundNoise

To achieve further realism in the simulations, background noise can be added to input signals. This function utilises already measured background noise signals from the semi-anechoic

chamber and adds them to the input signals. The functions demands a 16 microphone channels input in order to ensure that each microphone always has the background noise recording related to itself. The output from this function reflects the actual noise floor present in the semi-anechoic chamber.

3.3 Measurement Design

This section explain how measurements were designed, their purpose and the development process. The measurements aims to evaluate the beamformers real world performance in a similar fashion as it would be used by Volvo NVH engineers performing measurements in the semi-anechoic chamber.

3.3.1 Pulse Trains

A pulse train is a signal which contains pulses spaced out over time span with some silence between each pulse. In this context the pulses are sinusoidal tones at different frequencies. The purpose of using pulse train signals in this context is to estimate the individual beamformers frequency resolution at various positions within the possible source area in the semi-anechoic chamber. Figure 3.6 shows the setup during a pulse train measurement session at a specific location in the semi-anechoic Volvo Noise & Vibration Laboratory. The ULAs in the chamber is also visible on the right side of the picture.

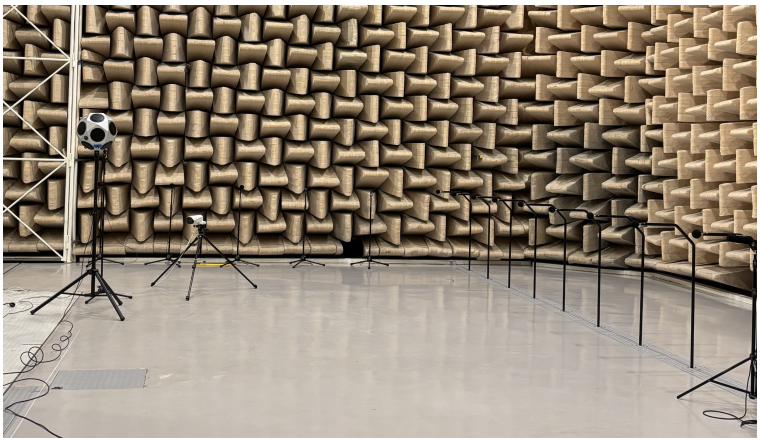


Figure 3.6: Measurement setup inside the semi-anechoic chamber during pulse train signal acquisition.

The pulse train contains 1/3-octave band center frequencies in the range 40 Hz to 2.5 kHz and in between each pulse there are some silence for spacing purposes and easy separation of each frequency. Compared to the frequency range stated in the aims and purposes for this master's thesis the pulse train signal contains 2 extra center frequencies, one below and one above.

Figure 3.7 depicts the pulse train signal in time domain and it contains both the voltage input signal to the source loudspeaker amplifier and the recorded sound pressure microphone signal closest to the source.

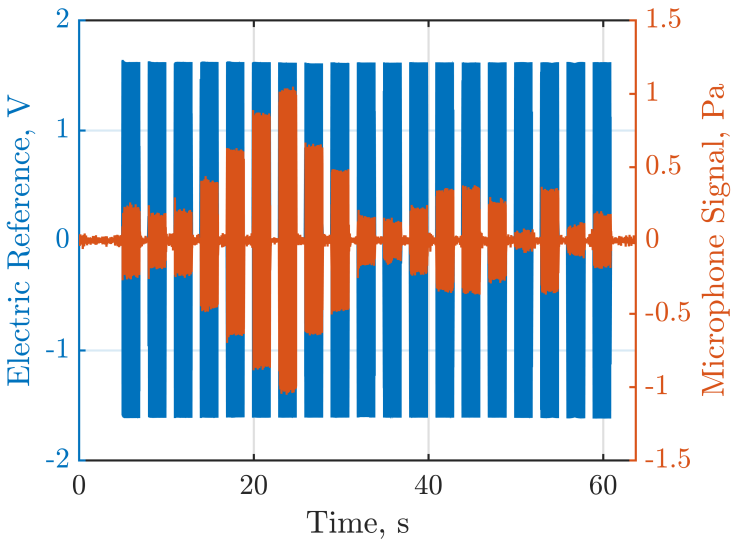


Figure 3.7: Pulse train time signal.

The source positions for the pulse train measurements were selected to cover all relevant array global bearing angles from 0° to a approximately −50° which is somewhat inflated in order to cover a larger area than a standard vehicle size. The angle also impacts expected grating lobe influence on a certain beamformers array pattern. Small angle positions is expected to have less grating lobe influence than large angle positions at the same frequency.

Signal-to-Noise Ratio

Different SNR cases were tested during two measurement sessions to estimate the beamformers robustness under varying noise levels. The initial plan was to use three different SNR cases, high with no additional noise. Case medium with a small amount of additional noise and case low with more additional noise. The extra noise was added to the sound field using an additional noise source, playing white noise and assumed to be part of background noise when calculating SNR like seen in equation 3.2.

During the first of the two measurement sessions we experienced distortion artefacts and erroneous sound level behaviour at different frequencies. Examining the signal path the source loudspeaker was deemed to be the culprit and also the only part of the signal chain which was non-replaceable. To mitigate these problems the SNR cases had to be updated. Trying to avoid the distortion peaks, the amplifier was set to a low enough level to minimize their contribution, this reflects the **medium** SNR case. The **low** SNR has the external noise source added. The **high** SNR case is measured with increased output level from the malfunctioning loudspeaker with no care taken to reduce distortion artefacts, this case was recorded during the second measurement session. The erroneous levels can also be see in the recorded microphone signal visible in figure 3.7 where the difference between both curves highlights the loudspeaker and amplifiers uneven level response over the frequency spectrum.

$$\mathrm{SNR} = 20\log_{10}\left(\frac{\tilde{P}_{\mathrm{pulse}}}{\tilde{P}_{\mathrm{background}}}\right) \tag{3.2}$$

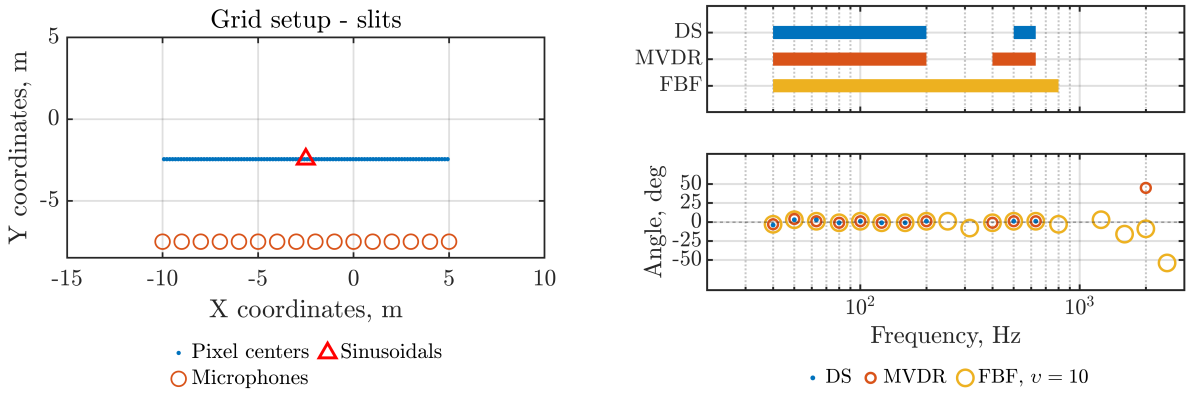
The SNR for pulse trains is always calculated and presented per 1/3-octave band, both background and pulse signals are filtered using the same 1/3-octave band bandpass filter. Using standard, non-filtered Root-Mean-Square (RMS) values would underestimate the SNR, due to varying noise content over a wider frequency range. The 1/3-octave band filtering is also robust against spectral leakage, which could be present in other SNR methods based on averaging, smearing the sinusoidal pulse. Instantaneous estimation from a magnitude spectrum or similar could suffer from missing frequency content and make non valid SNR estimates.

Beamformer Evaluation

The pulse train measurements are used to evaluate the valid frequency range for source localization purposes for each implemented beamformer technique. The true source angle can be compared to the individual beamformer outputs and the result can be compared to estimate their performance in realistic conditions.

With the recorded microphone signals and the true source position from the measurements, each beamformers performance can be evaluated. Using an external function called `gridBuilder` which creates beamforming locations at a certain distance from the array, spatial data for the beamformers input is created. This function is described in previous subsection 3.2.1.

Figure 3.8 depicts the beamformer evaluation output and grid setup from `gridBuilder`. This data comes from real measurements, this specific setup has been omitted from further analysis in favour of other measurements.



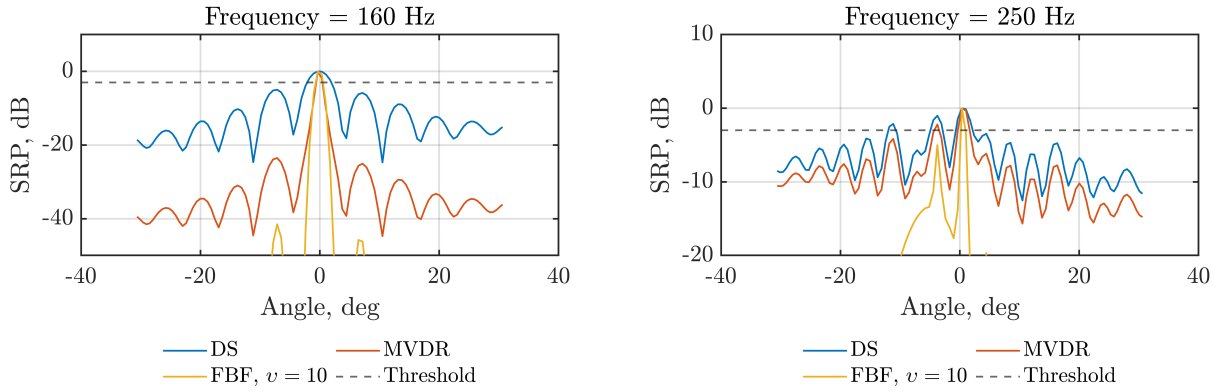
(a) Grid setup. **(b)** Beamformer evaluation.

Figure 3.8: Pulse train beamformer evaluation.

In order to interpret these figures correctly, it is important to have some pre-knowledge. The lower right plot in figure 3.8 illustrates estimated source angle, true source angle per 1/3-octave band center frequency. Each circles shows a specific beamformers source location at that specific frequency.

What the script essentially does is evaluating the beamformers produced SRP based on two criterion, its ability locate the source at the correct angle within a certain tolerance value, usually set to 10°. At some frequencies, certain beamformers symbols is missing, like DS at 250 Hz and 315 Hz. At these frequencies the array pattern was deemed to degraded to do any meaningful source localization, usually the array pattern has more than one peak higher than another tolerance value (usually −3 dB from normalized 0 dB) and its quality is deemed to be insufficient. See figure 3.9 showing a non-degraded SRP for 160 Hz and a SRP at 250 Hz, where both DS and MVDR has more than one peak above the −3 dB threshold, they are missing circles in the lower right corner plot of figure 3.8.

The upper right corner plot of figure 3.8 estimates valid frequency range for each beamformer based on the data which passed both tolerance values checks, with an extra addition. Two or more correct angle estimation in a row needs to be achieved, this excludes one frequency hits if both directly lower and higher frequencies are incorrectly estimated. Just a singular frequency hit is not deemed to be enough to sufficiently state that it is in valid frequency range.



(a) 160 Hz non-degraded.

(b) 250 Hz, degraded.

Figure 3.9: Pulse train beamformer evaluation.

3.3.2 Noise Signals

Noise signals was recorded at several positions along the $y = 0$ m line within the semi-anechoic chamber. The measurement consisted of a loudspeaker playing white noise at the same physical height as the microphones. The measurements served dual purposes, an initial IR estimation trial via random noise excitation and $H1$-estimator. Also, using the same excitation signals, confirming if the expected comb pattern would appear in a frequency domain representation of the recorded signals due to constructive/destructive interference from the floor reflection at the individual microphone locations. Figure 3.10 shows the measurement setup used during the noise recordings.

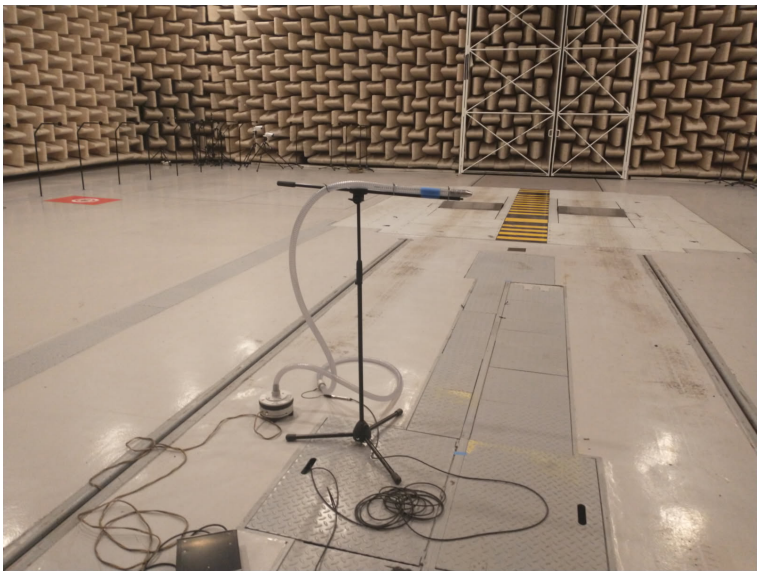


Figure 3.10: Noise measurement setup along the $y = 0$ m line in the semi-anechoic chamber.

After the session we had problems extracting valid IRs from the data and we realized that due to frequency range constraints of the loudspeaker, the frequency content of the IRs would not be sufficient. The loudspeaker lower frequency range extended to 100 Hz, while the timbre of a truck contains large amount of energy below this frequency, this significantly reduced the usefulness of the recorded signals for IR purposes.

Up to this point, both IR and ISM had been developed side-by-side as possible candidates for developing floor reflection simulation environments. As a consequence of these measurements, and exhausting available loudspeaker options, ISM was chosen as the foundation for developing a reflection simulation environment. ISM had the advantage that it would no be dependent on any input measurements in order to create simulation environments, the result is also not dependent on a spatial room position, as IR simulations would be. The ISM simulation behaviour would have to be verified with measurements for the frequency ranges and positions where valid data exists. The implemented ISM environment is further introduced in subsection 3.2.2. The noise measurements still proved useful as they confirmed that the expected comb pattern due to the floor reflection was visible in the data. They were used to validate the ISM floor reflection simulation behaviour. Figure 3.5 illustrates measured and simulated noise signals and their PSDs. Figure 8.6 lower part of plots, presents different noise source measurement, the instant signal, its related PSD and simulated PSD.

3.4 Validation Metric Design

For evaluation and comparison purposes of the different implemented techniques, a set of validation metrics has been established. These metrics are designed to assess the accuracy, robustness and spatial resolution of each method. All techniques are evaluated under identical conditions using the defined metrics to maintain consistency across performance assessment. A quantitative performance analysis was conducted under stochastic condition to evaluate

each technique. Both the microphone offsets and sensitivity variation metrics are evaluated using Monte Carlo simulations which uses repeated random sampling of variables to model the probability of different outcomes. For each of these tests, 100 independent trials were performed where random offsets are introduced within a set of predefined bounds. The results from all trials are aggregated to estimate the average and variability in performance.

The simulations use generated tonal signals at 1/3-octave band center frequencies ranging from 40 Hz to 2500 Hz. To evaluate directional sensitivity the source is positioned both on-axis and off-axis relative to the array broadside direction as shown in Figure 3.11. The measurement configuration is designed to suit each technique while maintaining reproducibility and standardization, enabling consistent comparison across techniques at a later stage.

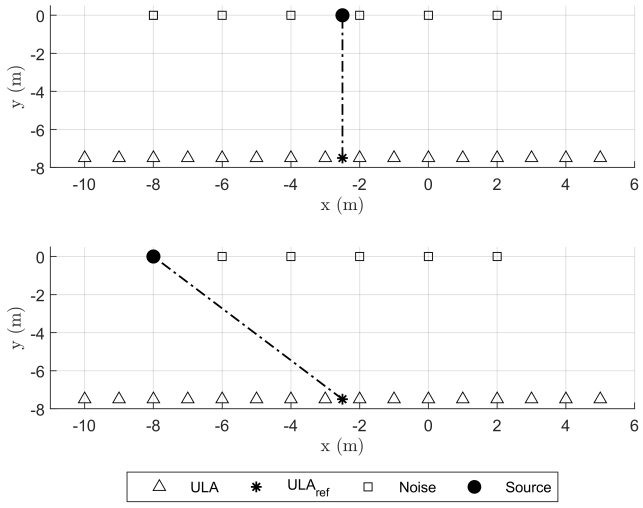


Figure 3.11: Simulated source configurations used for validation metric evaluation. (Top) On-axis setup with the source aligned with ULA_{ref} (DOA = 0°). (Bottom) Off-axis setup with the source positioned at an angle of DOA = -36° relative to ULA_{ref}.

3.4.1 Microphone Position Offsets

This metric evaluates how sensitive individual beamforming techniques are to inaccuracies in the microphone positioning. It tests the robustness of the method when the actual placement deviates from the assumed reference geometry. In each trial, random positional offsets are applied independently to every microphone within a maximum range of ±5 cm in x-y coordinate system. This simulates practical uncertainties in array geometry and mounting such as slight shifts or misalignments in microphone positions. The offsets are distributed within the specified bounds ensuring that the Monte Carlo simulation covers a wide range of realistic offsets as shown in Figure 3.12.

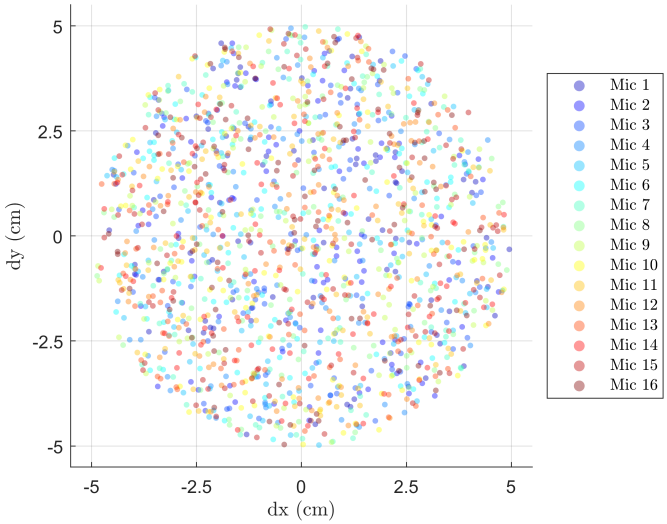


Figure 3.12: Microphone position offset distribution

The impact of these offsets are assessed by analysing the localization error across all trials. For each offset configuration the technique being evaluated attempts to localize a known source and the deviation from the true source position is stored. The average localization error and its standard deviation are used as statistical indicators of the specific techniques robustness to microphone placement inaccuracies. This provides a general performance overview by summarizing the localization results across all trials for each source position. In addition to localization performance the array response is also evaluated to asses any distortion in array patterns. Once the Monte Carlo simulation is completed under noiseless conditions further tests are conducted by introducing increased levels of noise into the system.

3.4.2 Microphone Sensitivity Variations

Microphone sensitivity mismatches are common issue in array signal processing due to manufacturing tolerances, environmental influences or calibration errors. This metric examines how individual microphone gain differences affect each techniques performance. The goal is to determine how sensitive they are to variations in microphone gain and sensitivity.

In this test each microphone is assigned a random gain offset from a uniform distribution. The nominal gain is set to 0 dB and variations are introduced independently for each microphone. This setup simulates realistic conditions in which microphones sensitivity deviate slightly from their ideal case. As in the microphone offset tests, a Monte Carlo simulation is performed and the array response is used to evaluate the specific techniques performance under gain variability.

3.4.3 Spatial Resolution

This metric measures individual techniques ability to distinguish between two closely spaced sources. It provides an objective view of the spatial focusing performance and angular resolution.

Figure 3.13 show the simulation setup used for this evaluation. Two identical sources are initially positioned on opposite sides of the array. For each iteration the sources are incrementally moved closer together and the array response is computed. This process is then repeated until the two sources fuses together, becoming indistinguishable from one another in the array response and produces one mainlobe containing two sources. The spatial resolution is then determined as the smallest source separation distance at which a technique is able to resolve two different sources.

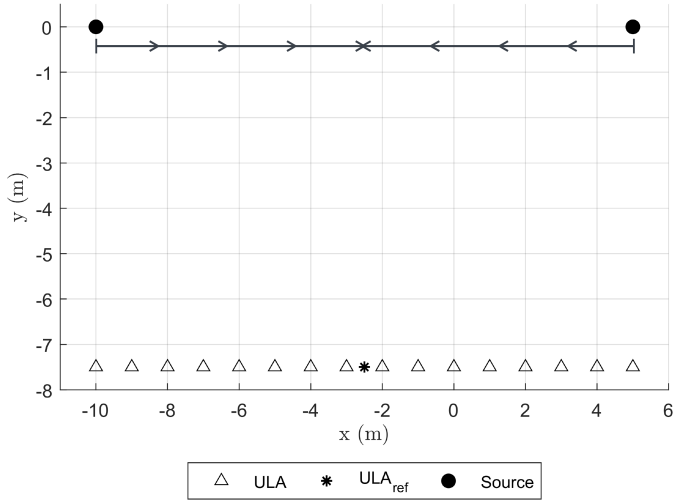


Figure 3.13: Simulation setup for spatial resolution evaluation.

4

Delay-and-Sum Beamforming

This is the first of four technique chapters and it presents the first technique selected for evaluation, the Delay and Sum (DS) beamformer. The chapter contains a description of the technique, the theoretical background specific for it, details about the specific DS scripts design and implementation. It covers the validation metrics simulated results. All data gathered is then used to evaluate the DS beamformers performance in the analysis subsection. The result should highlight its strengths and identify aspects which could form a basis for selecting the next technique to implement.

4.1 Technique Description

The DS beamforming technique is a fundamental algorithm for array signal processing and accurate localization of acoustical sources. It is a simple and straightforward technique based on time-aligning signals from individual microphones which steers the arrays response towards a certain direction, amplifying signals from the desired direction and attenuating from others.

Due to the techniques simplicity, its intuitive structure, and foundational role in beamforming technology, the DS beamformer is well-suited as a baseline model in the acoustic camera software. It serves as a valuable reference for evaluating the performance of more advanced localization techniques and serves as an intuitive link between microphone array geometry and source direction.

4.2 Theoretical Background

The fundamental concept of DS beamforming is based on time-aligning signals recorded by a microphone array. As a wavefront propagates across an array, each microphone captures a time delayed version of the original signal. By applying compensating delays that time-align the signals originating from a specific direction and then summing them, the array effectively reinforce the desired signal content. These compensating delays is directly related to the time it takes for a signal to propagate between microphones [17].

4.2.1 Time Delay

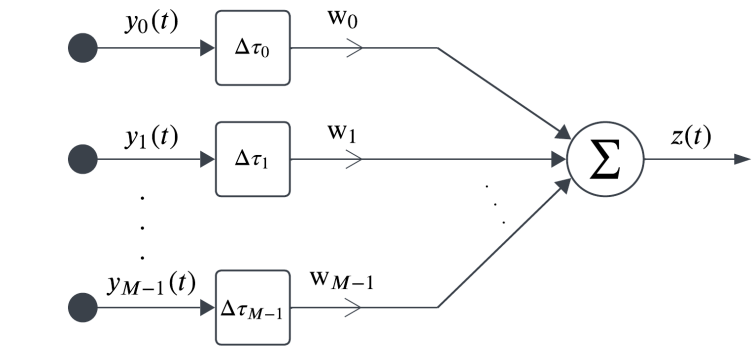
Consider an array of M sensor elements, in the case of acoustic signals, the sensor elements are microphones. Each microphone part of the array samples the wavefield spatially at the geometrical location m_k, see figure 2.6 for a visual representation of the setup. The DS algorithm then applies a delay $\Delta\tau_{mk}$ and an amplitude weight w_m to the output of each

microphone. In the standard DS beamformer, uniform amplitude weights are applied across the array, i.e $w_m = 1$ for all microphones.

The calculated time delays $\Delta\tau_{m_k}$ is applied to each microphone signal and summed up to form the beamformed output. This summation results in constructive or destructive interference depending on the bearing angle θ_c. When the delayed signals align from a particular direction they add up coherently, amplifying the summed source signal. Simultaneously signals from other directions result in large misalignment in phase, leading to attenuation. The DS beamformer's output is defined as:

$$z(t) = \sum_{m=0}^{M-1} w_m y_m(t - \Delta\tau_{mk}) \tag{4.1}$$

As a result the array response approximates the source location by scanning over different bearing angles (θ_c) and distances. Identifying the direction that maximizes the summed signal. This forms the foundation of DS beamforing based source localization. Figure 4.1 visualizes the DS beamformer. Where microphone signals $y_m(t)$ is delayed, weighted and summed.



a

Figure 4.1: Delay-and-Sum beamformer process

To accurately steer the beamformer, the time delays τ_{m_k} must be calculated based on the assumed shape of the incoming wavefront. In the general case of a spherical wavefront, in near-field scenarios, the delay at each microphone is given by:

$$\tau_{m_k} = \frac{r_{m_k} - r_c}{c} \tag{4.2a}$$

where r_c is the distance from the array center to the focus position, c is the speed of sound, and r_{m_k} is the distance from the focus position to the m_kth microphone, defined as:

$$r_{m_k} = \sqrt{r_c^2 + d_{m_k}^2 - 2r_c d_{m_k} \cos\theta_c}$$

This follows the law of cosines, with θ_c being the angle between the array's perpendicular axis and the direction to the focus point. The variable d_{m_k} denotes the position of the m_kth microphone relative to the center of the array and is given by:

$$d_{m_k} = \left(m_k - \frac{M-1}{2}\right) l$$

Replacing r_{m_k} in Equation 4.2a with its definition and d_{m_k} above leads to the following expression for the time delay τ_{mk}:

$$\tau_{m_k}(\theta_c) = \frac{1}{c}\left[\sqrt{r_c^2 + \left(\left(m_k - \frac{M-1}{2}\right)l\right)^2 - 2r_c\left(\left(m_k - \frac{M-1}{2}\right)l\right)\cos\theta_c} - r_c\right] \tag{4.2b}$$

In the far-field where r_c is large compared to the array aperture, the spherical wavefront approximates a plane wavefront. Under this condition Equation 4.2b simplifies to:

$$\tau_{p_{m_k}}(\theta_c) \approx -\frac{1}{c}\left(m_k - \frac{M-1}{2}\right)l\cos\theta_c \tag{4.3}$$

The transition from spherical to planar wavefront assumptions illustrates the distinction between near-field and far-field beamforming. In the far-field, time delays vary linearly with microphone index, as shown in Equation 4.3, and the wavefront can be approximated as planar. In contrast, near-field conditions require spherical modelling due to the wavefront curvature, which introduces additional differences in the time delays across the array. Properly accounting for these differences is essential for accurate DOA estimation and effective beamforming performance.

4.2.2 Relative Time Delay Compensation

Figure 4.2 illustrates the core principle of DS beamforming. The figure shows four microphones focusing on a single source and highlights how each microphone signal is delayed to compensate for propagation differences before summation. Although the figure assumes a spherical model, its main purpose is to visualize the delay compensation process used in DS beamforming to achieve constructive interference at the focus point.

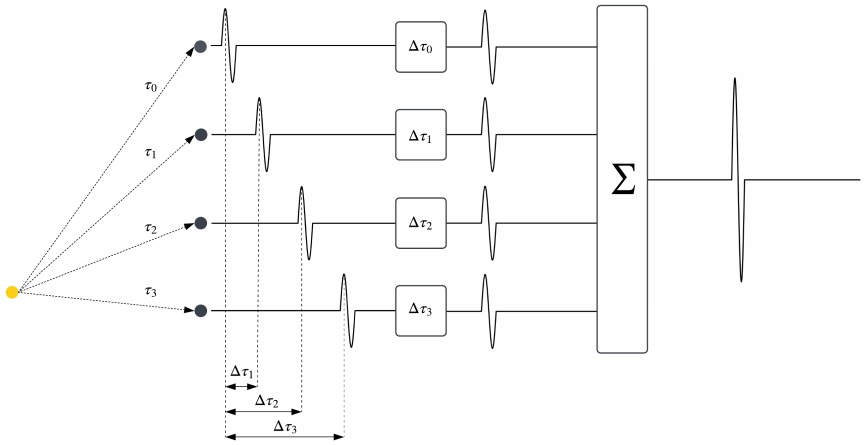


Figure 4.2: Schematic illustration of the delay and sum beamformer. The array is steered toward a specific source location by applying compensating time delays to the signals received at each microphone. These delays align the signals in time, allowing constructive summation of the wavefront from the target direction.

In the practical implementation of DS beamforming, the delays τ_{mk} defined in Equation 4.2 represent the absolute propagation times from a given focus point to each microphone in

the array. However, the DS beamforming process does not operate based on these absolute arrival times. Instead, the beamformer aligns the microphone signals by compensating for their relative timing differences, allowing constructive interference from a specific direction. To achieve this, delays are computed relative to the microphone that receives the signal last. That is the microphone for which the propagation delay τ_{mk} is maximal. The compensating delay applied to each microphone is given by:

$$\Delta\tau_{mk} = \max_{m} (\tau_{mk}) - \tau_{mk} \tag{4.4}$$

DS beamforming relies on accurately modelling the wavefront propagation and compensating the relative delays across the array. By aligning signals based on timing differences between microphones the array achieves constructive interference from a chosen direction. This theoretical foundation enables practical implementation through delay compensation for effective source localization.

4.2.3 Steering Vectors

In the frequency domain, delay compensation in beamforming is expressed using a steering vector. This vector describes the phase shifts corresponding to the propagation delays between a specified focus point in the scanning plane and each microphone in the array. For the mth microphone the frequency domain steering vector is given by:

$$a_m(\theta) = e^{-j\omega\Delta\tau_{mk}} \tag{4.5}$$

These complex exponential weights apply frequency dependent phase shifts to align the signals received at each microphone. When the assumed directions θ_c matches the actual source direction, the phase aligned signals sum coherently across the array, reinforcing the signal power from that direction. This allows the beamformer to steer its sensitivity spatially, enabling directional focus and improved signal enhancement. This frequency domain formulation of directional sensitivity acts as a foundation for more advanced beamforming methods such as the Minimum Variance Distortionless Response (MVDR) beamformer, enabling improved spatial selectivity and reduced interference from unwanted directions.

4.2.4 Fractional Delay Filter

Previous subsection 4.2.1 introduces the propagation delay τ_{mk} and how it is computed for each microphone based on the wavefront geometry and focus point. Next subsection 4.2.2 introduced the concept of relative delays $\Delta\tau_{mk}$, which are used to time-align the microphone signals to achieve constructive interference from a desired direction. In practice, these compensating delays $\Delta\tau_{mk}$ are generally non-integer multiples of the sampling period. Since discrete-time systems can only shift signals by whole samples, fractional delays must be implemented using digital filters. This approach follows the method proposed by Välimäki and Laakso [18] where fractional delay filters are used to implement sub-sample alignment via truncated and windowed sinc functions. This is achieved by designing fractional delay Finite Impulse Response (FIR) filters which provide sub-sample precision by shaping the filter phase response appropriately. Each microphone signal $y_m(t)$ is a time delayed version of the original source signal $x(t)$ due to the propagation time:

$$y_m(t) = x(t - \tau_{mk}) \tag{4.6}$$

Time-aligning all microphone elements, their relative delay at each microphone needs to be compensated, as in Equation 4.4. This principle is used to design the fractional delay filters. The ideal frequency response of a delay filter is defined as:

$$H(\omega) = e^{-j\omega D} \tag{4.7}$$

with the corresponding magnitude and phase characteristics:

$$|H(\omega)| = 1$$
$$\angle H(\omega) = -\omega D$$

where $D = \Delta\tau_{mk} \cdot f_s$ is the desired delay in samples. This filter is an linear phase all-pass filter, meaning it ideally introduces no amplitude distortion and applies a constant delay D across all frequencies. Which is essential for keeping the shape of the signal while achieving sub-sample alignment. Taking the inverse DTFT gives the corresponding impulse response:

$$h(n) = \mathcal{F}^{-1}\{H(\omega)\} \Rightarrow h(n) = \frac{1}{2\pi}\int_{-\pi}^{\pi} e^{j\omega(n-D)}d\omega \Rightarrow$$
$$\Rightarrow h(n) = \frac{sin[\pi(n-D)]}{\pi(n-D)} \Rightarrow h(n) = \text{sinc}(n-D) \tag{4.8}$$

This shows that the ideal impulse response of a fractional delay filter is a shifted sinc function. In time domain the delayed signal $\tilde{y}_m(n)$ is obtained by convolving the input $y_m(n)$ with the impulse response of the fractional delay filter $h(n)$:

$$\tilde{y}_m(n) = y_m(n) * h(n) \tag{4.9}$$

To illustrate the effect of fractional delays Figure 4.3 shows two sinc functions representing delay filter impulse responses. In the top plot the delay D is an integer value and the sinc function is centered exactly on a sampling point. As a result only one sample has non-zero value while all other coincide with the zero-crossings of the sinc [18]. In contrast the bottom subplot shows the case where D is a non integer delay. The sinc function is shifted off the sampling grid and all samples are non-zero values. This illustrates how fractional delay filters spread the energy across multiple taps to achieve sub-sample shifts.

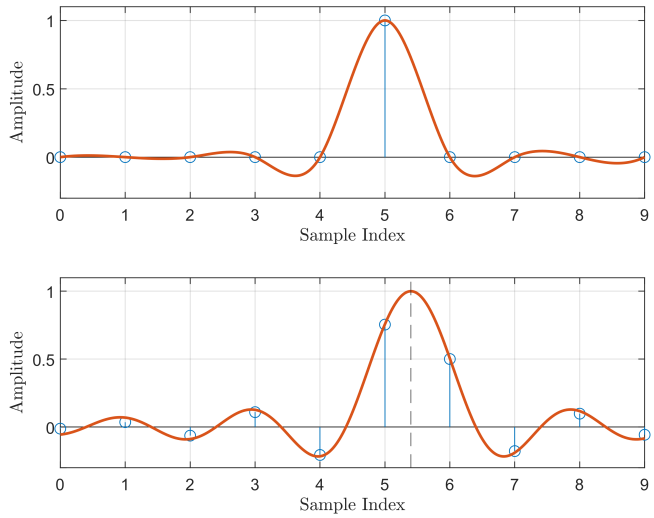


Figure 4.3: Comparison of FIR filter coefficients for an integer delay and fractional delay, following the concept in [18]. (Top) sinc function (line) shifted by $D = 5.0$ samples, (Bottom) sinc function (line) shifted by $D = 5.4$ samples, circles show the sampled sinc function values, dashed lines indicate the center point of the shifted sinc function.

4.3 Implementation

The implementation of the DS beamformer algorithm is constructed as a nested function where the outer function handles spatial data and calculates geometrical distances while the inner function calculates fractional delays. Inside the nested function, steering vectors are calculated and remodelled as filter IRs and applied to input time domain signals. Figure 4.4 illustrates the DS function as a flowchart where the nested function is marked as a dashed rectangle.

As **Input** the function takes time domain microphone signals, position data relating to the ULAs, also angles and distances for the beamformer to aim at. Based on the available geometrical data the main function calculates the **Relative Time Delay** which is divided into two steps, calculating the time distance it takes for a sound wave to travel from a certain aiming location to each microphone and then scaling them based on the microphone furthest from the source, the microphone which took the longest time to receive a signal. This time difference scaling of the microphone signals is the relative part and the purpose is that the beamformer waits until the position in time where each microphone has been reached by a signal, then it calculates steering vectors on this inter-microphone delay.

The Fractional Delay nested function contains four steps in total, where the last step just applies the calculated delay filters to microphone signals. As input it uses the relative time delays $\Delta\tau_{m_k}$ and calculates the **Steering Vectors** as equation 4.5 suggests. The equation utilises angle frequency ω and the number of frequencies is based on $\Delta\tau_{m_k}$ but recalculated into samples, the resulting frequency vector represents a single-sided frequency response and when inputted into the steering vectors equation the resulting steering vectors can be viewed

as single-sided frequency responses of a steering vector filter. Next step, **Filter Frequency Response** recalculates the single-sided frequency steering vector responses into double-sided responses and by applying an **IFFT** to the double-sided responses, the result are steering vector IRs, which can be applied to the input signal. The steering vector filters therefore contains a certain microphone array setups timing response towards a certain aiming location which is truly broadband.

The **Convolution** step applies the steering vector filters to the input signals in time domain and the result is the **Output** of the beamformer, a matrix containing all individually beamformed signals and summation of them which is represented as a $\sum$ in figure 4.2.

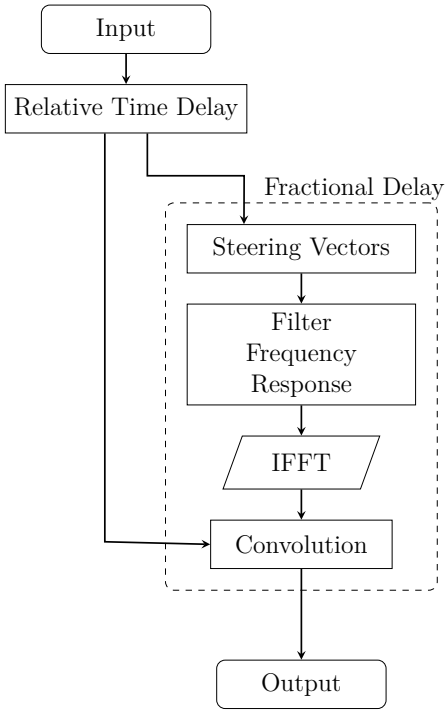


Figure 4.4: DS beamformer flowchart.

4.4 Validation Metrics

This section presents the validation results for the DS beamformer using the validation metrics explained in Chapter 3.4. The following results highlight the DS beamformers performance under each evaluation scenario. Results provided is to demonstrate the DS beamformer response to microphone offsets, sensitivity variations, and source separability.

4.4.1 Robustness - Microphone Placements

Evaluating the robustness of the DS beamformer due to microphone placement errors, a 100 trial Monte Carlo simulations is performed under two different acoustic conditions. The first condition represents an ideal scenario with no added noise or ground reflections while the second includes both additive white Gaussian noise and background noise recorded in the

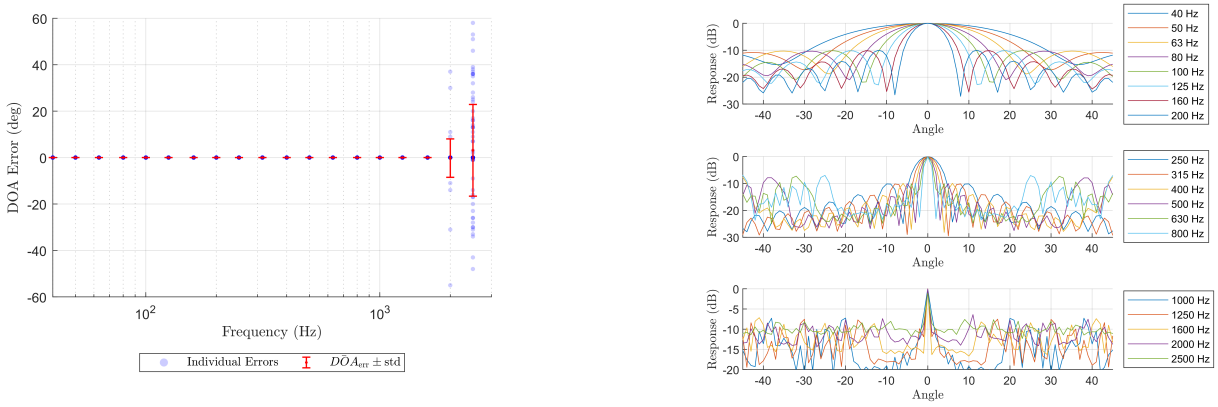
semi-anechoic chamber. Each of these scenarios is tested for both on-axis and off-axis source positions too understand how positional errors influence directionality.

Ideal Case - on-axis

In the absence of noise the DS beamformer performance reflects the isolated impact of only the microphone position offsets with the source place on-axis (DOA = 0°). For each simulation trial a random spatial offset is randomly applied to all microphone positions within a ±5 cm range in the x-y plane, simulating realistic array mounting tolerances.

The distribution of DOA estimation errors across all trials is presented in Figure 4.5a, where the error is computed as $\mathrm{DOA}_{\mathrm{error}} = \mathrm{DOA}_{\mathrm{ref}} - \mathrm{DOA}_{\mathrm{offset}}$. Each individual trial is plotted alongside the mean and standard deviation for every frequency band. The results show that for frequencies below approximately 2000 Hz the DOA estimates are clustered around the true direction. Indicating high robustness to positional offsets. Starting from 2000 Hz and upward a increased variability emerges in the DOA estimates, suggesting that the DS beamformer becomes more sensitive to geometric inaccuracies at higher frequencies due to the shorter wavelengths.

The corresponding averaged array responses are shown in Figure 4.5b, which represents the mean spatial response computed across all Monte Carlo trials using only the offset positions. This captures the typical beamforming performance under position offset conditions while excluding the ideal geometry. The figure highlights how the mainlobe remains focused near the true source direction despite microphone offsets.



(a) Distribution of estimated DOA

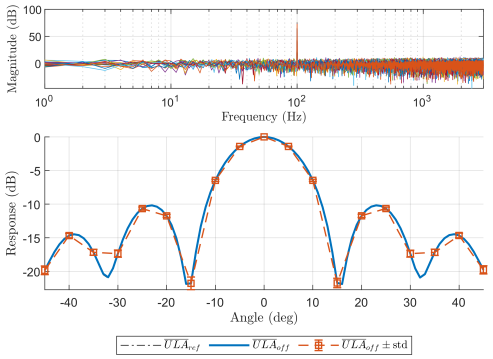
(b) Average Array Responses

Figure 4.5: Performance of the Delay-and-Sum beamformer in the noiseless baseline scenario when source is placed on-axis. (a) illustrates the distribution of DOA estimates across all Monte Carlo trials including the average estimated DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.

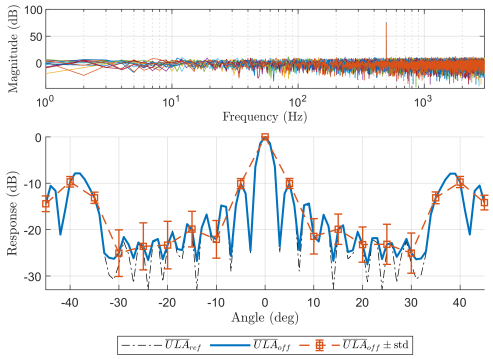
While higher DOA errors at higher frequencies are observed the overall response structure remains coherent, meaning that the majority of DOA estimations is still at the true direction. Looking at the average array responses in figure 4.5b, note that at 400 Hz grating

lobes begin to appear within the scanned angular range. As frequency increases, the grating lobes become more pronounced and progressively approach the mainlobe due to reduced wavelength relative to microphone spacing [9]. This confirms that the DS beamformer exhibits a strong degree of robustness to moderate microphone misplacements in the idealized on-axis case, especially at lower frequencies where spatial aliasing effects are less pronounced.

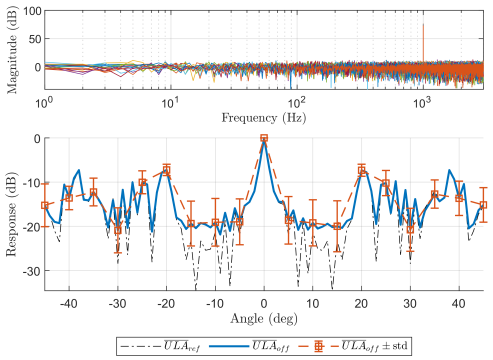
To further illustrate the frequency dependent impact of microphone placement errors Figure 4.6 presents the array response variability across four representative frequencies; 100 Hz (low), 500 Hz (mid), 1000 Hz (high) and 2000 Hz (high). These frequencies are selected to provide a concise yet representative overview of the beamfomer's behaviour across the spectrum rather than displaying results for all available frequency bands. For each sub-figure the top panel shows the Frequency Response Function (FRF) of the input signal for all microphones of the array. While the bottom panel displays the averaged array response across all trials using the offset microphone positions. The dashed-line in combination with box plots indicates the standard deviation at fixed angular steps, quantifying the response variability across trials. The dash-dotted line represent the reference array response obtained using the ideal microphone configuration.



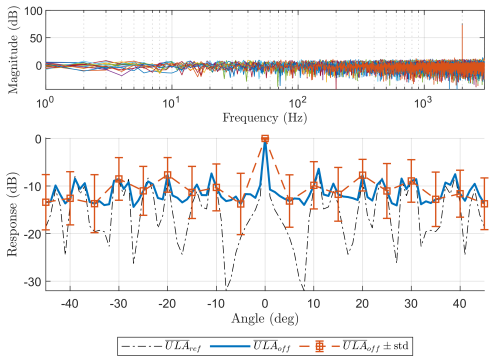
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz



(d) 2000 Hz

Figure 4.6: Simulated beamforming performance under microphone position misplacement at four frequencies. Each subplot shows the FRF (top) and the averaged array response with standard deviation (bottom) across all trials. Dash-dotted lines denote the reference array response using reference microphone positions.

At low frequencies (Figure 4.6a), the array response remains relatively stable with minimal variations across trials. This is reflected in the box plot by the whiskers being tightly clustered around the mean, indicating low spread at the evaluated angular position of the array response across all trials. The response from the ideal microphone configuration is not visibly distinguishable, indicating that microphone placement offsets have negligible impact in the low frequency range.

In the mid frequency region, Figure 4.6b shows the array response for a 500 Hz source. At this frequency the array response begins to exhibit increased variability across trials. This is observed as a broader spread in the box plots where the whiskers extend farther from the mean, indicating a growing sensitivity to microphone placements. Notably the whiskers around the mainlobe remain tightly clustered, suggesting that the beamformer continues to accurately localize the DOA despite introduced offset placements.

In this frequency range grating lobes begin to emerge within the scanning window while the mainlobe remains centred at the true source direction. The appearance of grating lobes

and increased spread imply that that spatial aliasing and geometric offsets are starting to influence the beamforming output more noticeably. When comparing the average array from the offset placements and the reference array a slight deviation is visible particularly in the side-lobe levels.

At the high frequency region, Figures 4.6c and 4.6d present the array responses for 1000 Hz and 2000 Hz sources respectively. At these frequencies the DS beamformer becomes increasingly sensitive to microphone misplacements. The box plots show a substantial increase in spread across trials, showing greater variability in the array response but the mainlobe remains clearly distinguishable in both cases. Indicating that the source is still predominately preserved.

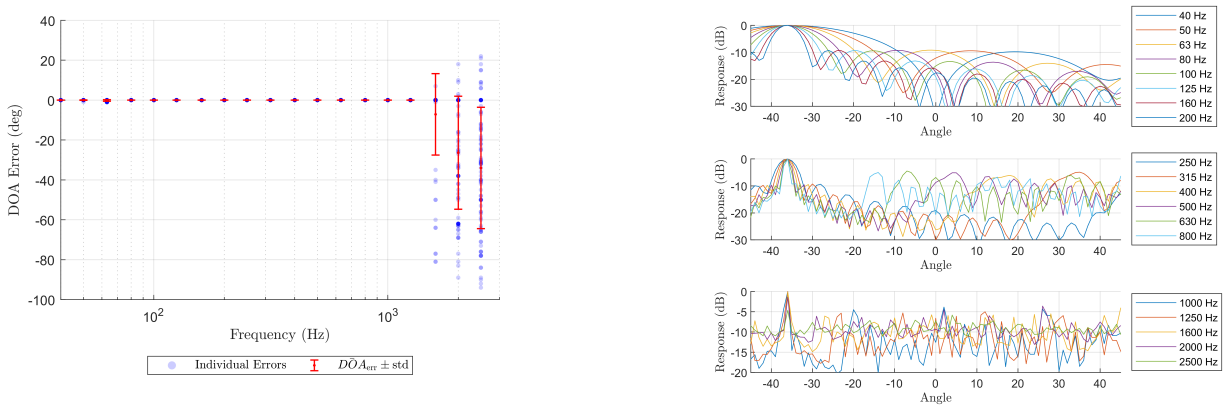
For the 2000 Hz source this observation is further seen by the DOA distribution in Figure 4.5a where only 8 out of 100 trials result in an incorrect DOA, corresponding to an error rate of 8% with a ± 5 cm misplacement. Despite the increased sensitivity at higher frequencies the DS beamformer still maintains a reasonable level of localization accuracy under moderate microphone offset placements.

Under idealized conditions microphone placement errors primarily affect the array response as the frequency of the source increases. At lower and mid frequencies the DS beamformer remains largely unaffected and consistently localizes the source.

Ideal Case - off-axis

While on-axis source placements provides a useful baseline for evaluating beamformer performance it represents a relatively rare and idealized scenario. In most practical applications sources are located off-axis resulting in asymmetric propagation paths to the array elements. To assess the beamformers performance under more realistic conditions a second idealized Monte Carlo simulation is conducted with the source positioned at an angle of $\theta = -36°$ relative to the array center. This falls within the effective scanning range of approximately $-45°$ to $45°$ where sources can be reliably localized given the array geometry and source distance.

Figure 4.7 presents the beamformer performance for the off-axis source placement. The DOA estimation results in Figure 4.7a show that the beamformer maintains accurate localization across most frequency bands similar to the on-axis case. However an additional frequency band exhibits a noticeable increase in estimation error. Recall that in the on-axis scenario DOA errors began to appear around 2000 Hz and 2500 Hz. In the off-axis case these two bands show even greater estimation error and the 1600 Hz band also displays increased deviation from the true source direction. Figure 4.7b shows that the mainlobe of the average array response is correctly shifted toward the true source direction across all frequencies. However the peak response is noticeably reduced for the highest frequency bands. Indicating a loss in beamformner gain response due to increased sensitivity to microphone position offsets.

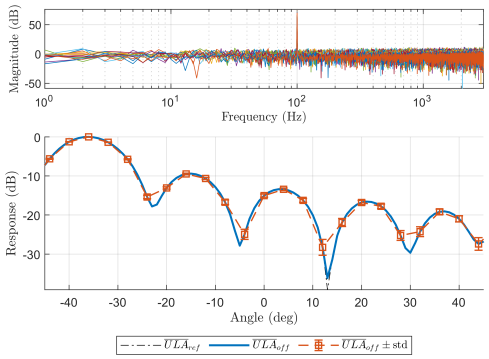


(a) Distribution of estimated DOA

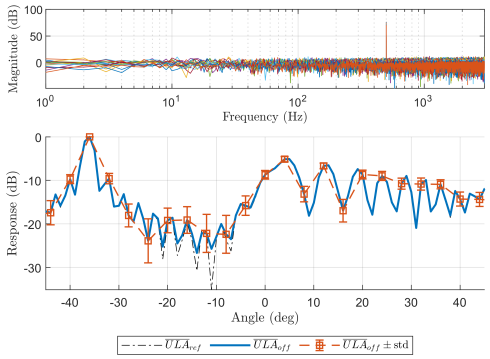
(b) Average Array Responses

Figure 4.7: Performance of the Delay-and-Sum beamformer in the noiseless baseline scenario when source is placed off-axis. (a) illustrates the distribution of DOA estimates across all Monte Carlo trials including the average estimated DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.

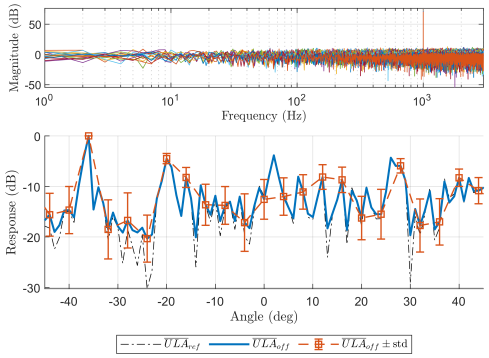
Figure 4.8 presents the array response variability. As expected the peak of the array response aligns with the true source direction around $\theta = -36°$ for all frequencies. Similar to the on-axis case the overall trend of increasing variability with frequency is clearly observed. At lower frequencies the standard deviation remains minimal and the reference and offset responses are nearly identical. As frequency increases variability grows and the average response begins to diverge from the reference. At 2000 Hz the boxplot reveal significantly larger whiskers at the mainlobe position. Indicating that many trials exhibited errors even at the estimated DOA. Meaning reduced localization at higher frequencies in the off-axis condition compared to the on-axis baseline.



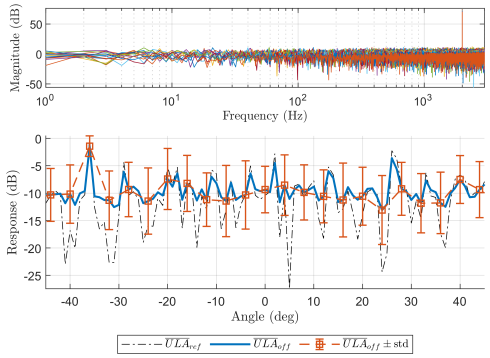
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz



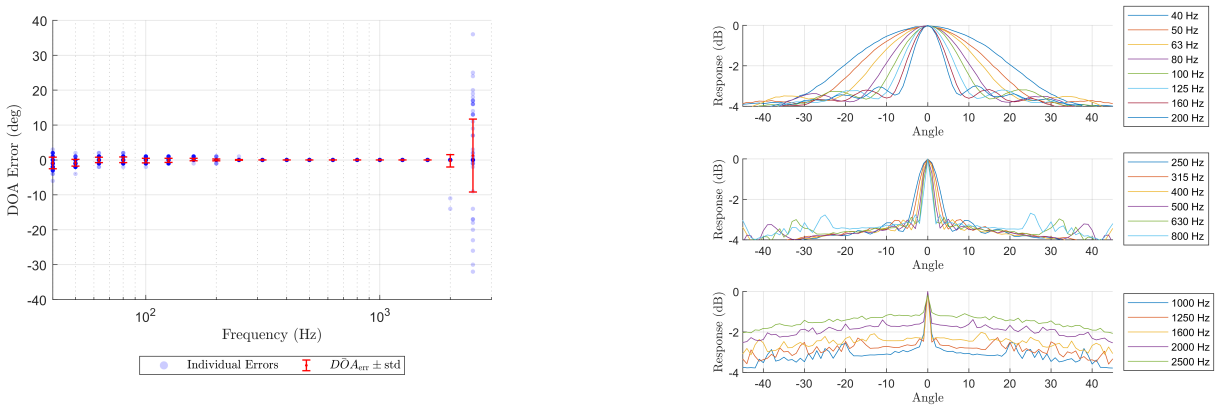
(d) 2000 Hz

Figure 4.8: Simulated beamforming performance under microphone position misplacement at four frequencies. Each subplot shows the FRF (top) and the averaged array response with standard deviation (bottom) across all trials. Dash-dotted lines denote the reference array response using reference microphone positions.

Added Background & Additive Noise Case - on-axis

Compared to the idealized case the presence of background and additive white noise is expected to degrade the spatial coherence between microphones. This can lead to increased estimation errors and more variability in the beamformer output.

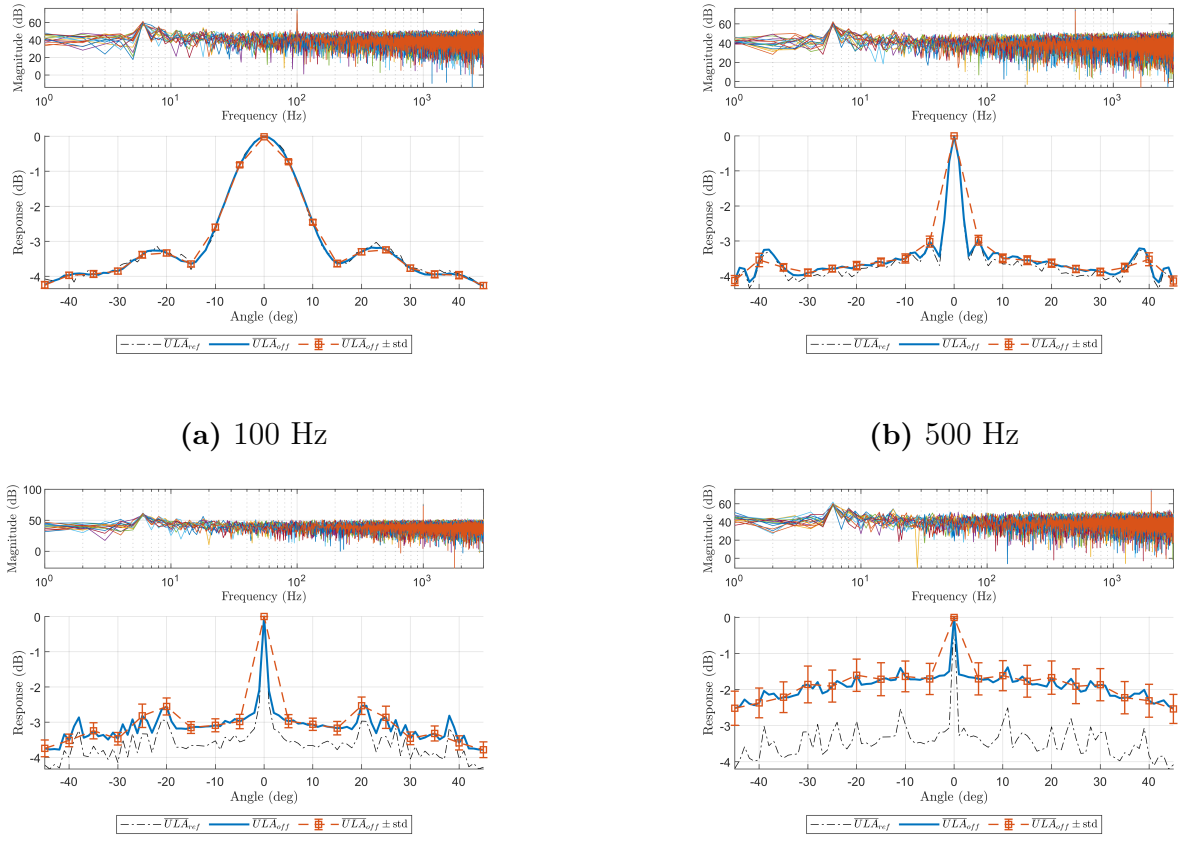
The DOA estimation error results in Figure 4.9a show that the low frequency region up to approximately 200 Hz exhibits noticeable but relatively small estimation offsets, typically within $\pm 5°$. In the frequency region between 250 Hz - 1600 Hz the beamformer maintains high localization accuracy. Implying that this is the most robust operating range under noisy conditions. At higher 2000 Hz the DOA spread has unexpectedly decreased compared to the idealized case at the same frequency. This could be the influence of the elevated noise floor that is masking the side and grating lobes as seen in Figure 4.9b. These results suggest that while the DS beamformer can still localize the source under noisy conditions its accuracy and resolution are significantly degraded over frequency ranges.



(a) Distribution of estimated DOA

(b) Average Array Responses

Figure 4.9: Performance of the DS beamformer in the presence of additive and background noise when the source is placed on-axis. (a) illustrates the distribution of DOA estimates across all trials including the average estimated DOA and its standard deviation. (b) presents the average array response patterns across all source frequencies under noisy conditions.



(a) 100 Hz

(b) 500 Hz

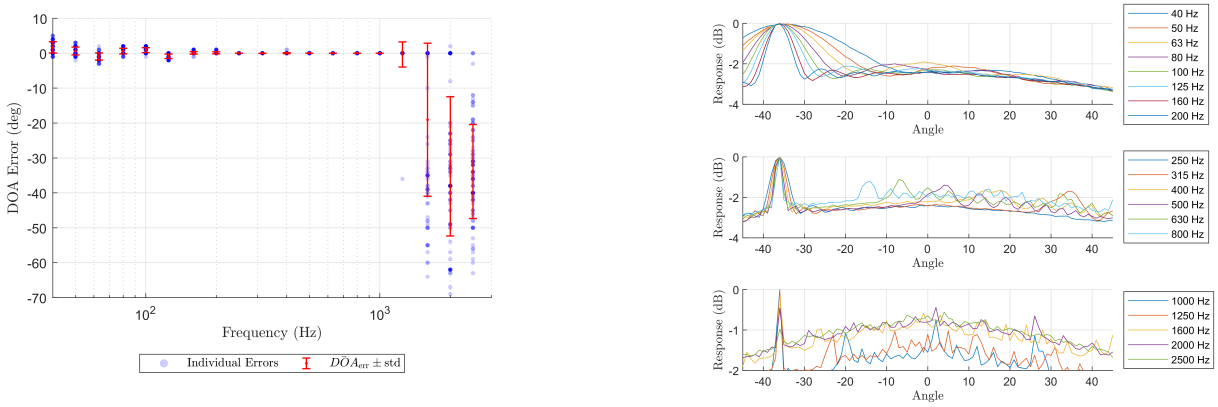
(c) 1000 Hz

(d) 2000 Hz

Figure 4.10: Simulated beamforming performance under microphone misplacement with added noise at four frequencies for an on-axis source.

Added Background & Additive Noise Case - off-axis

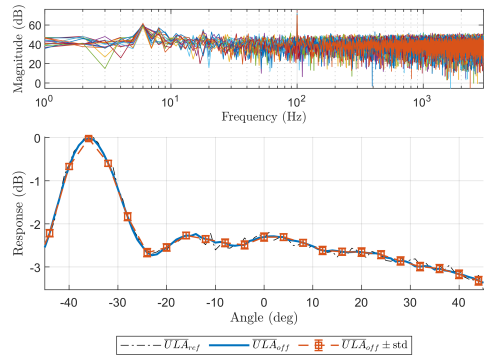
In the off-axis scenario under noisy conditions the DOA estimation error results in Figure 4.11a show a clear increase in localization errors. This is expected due to the combination of asymettric source placement and reduced spatial coherence. Despite this the DS beamformer maintains relatively stable performance across the frequency range of 125 Hz to 1000 Hz. Even at 1250 Hz only a signle error occurs across all trials, showing that this mid frequency range offers the highest resilience to noise and geometric offsets.



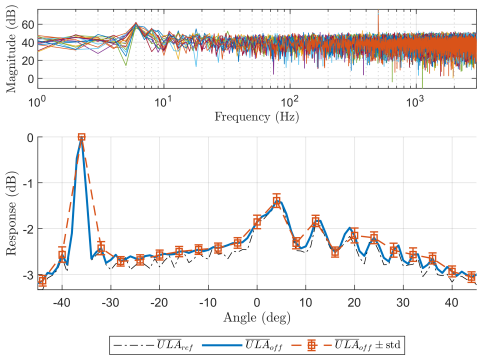
(a) Distribution of estimated DOA

(b) Average Array Responses

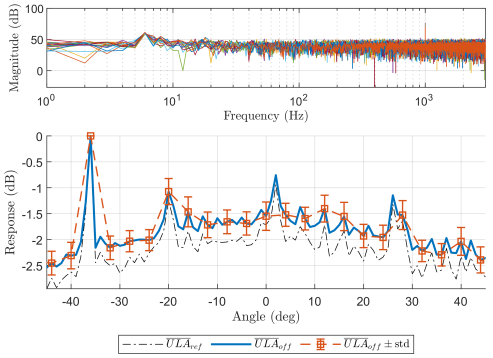
Figure 4.11: Performance of the DS beamformer in the presence of additive and background noise when the source is placed off-axis. (a) illustrates the distribution of DOA estimates across all Monte Carlo trials, including the average estimated DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies under noisy conditions.



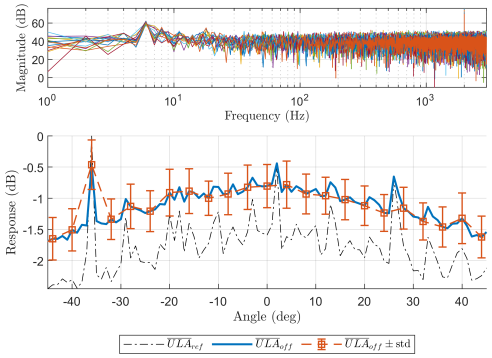
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz



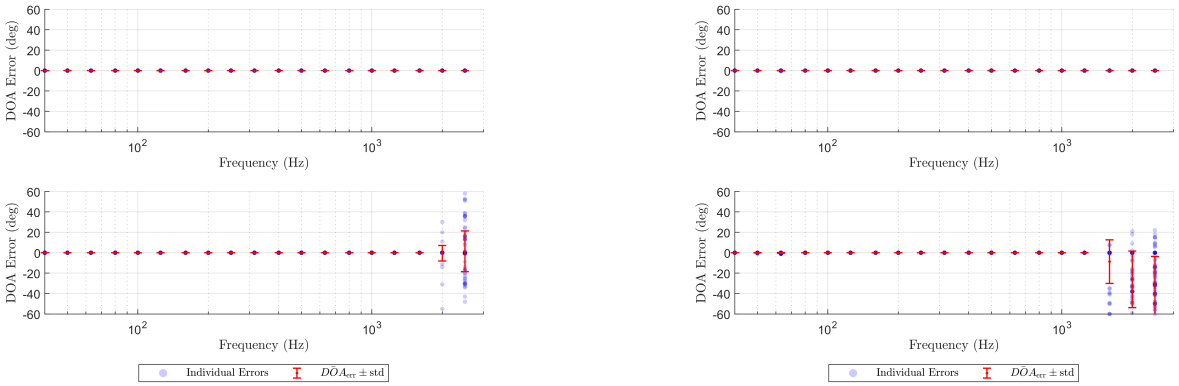
(d) 2000 Hz

Figure 4.12: Simulated beamforming performance under microphone misplacement with added noise at four frequencies for an off-axis source. Each subplot shows the frequency response function (FRF) (top) and the averaged array response with standard deviation (bottom) across all Monte Carlo trials. Dash-dotted lines indicate the reference array response using the ideal microphone positions.

4.4.2 Robustness - Varied Microphone Sensitivity

Evaluating the impact of varied microphone sensitivities, additional simulations were conducted using both reference and offset microphone positions. Since the DS beamformer operates mainly in the time domain by shifting discrete samples, the effect of sensitivity mismatches is expected to be minimal. The corresponding DOA estimation results are shown in Figure 4.13 for both on-axis and off-axis cases.

The DOA estimation results confirm this. For the reference geometry localization remains accurate across all frequencies. In the case with positional offsets and sensitivity variation combined the observed spread in DOA estimates begins around 1600 Hz for the off-axis case and 2000 Hz for the on-axis case. This aligns with earlier results and suggests that the observed variability is primarily driven by microphone placement errors, with sensitivity differences having negligible impact on DS beamforming performance.



(a) On-Axis **(b)** Off-Axis

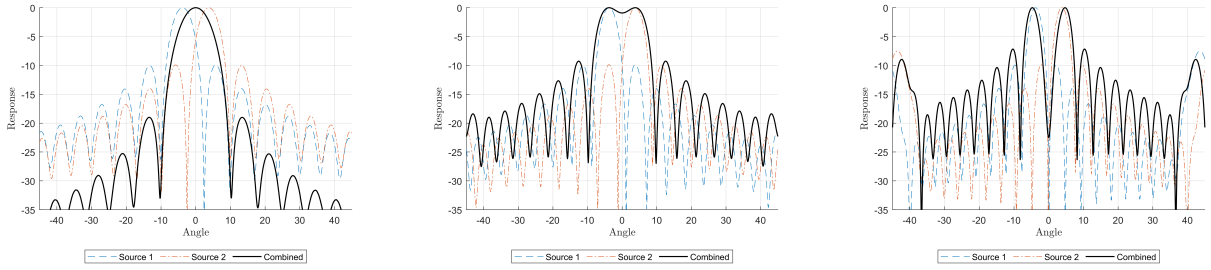
Figure 4.13: DOA estimation results under varied microphone sensitivity conditions for (a) on-axis and (b) off-axis source positions.

4.4.3 Spatial Resolution

The spatial resolution of a beamformer defines its ability to distinguish and separate multiple acoustic sources. One definition of resolution is the Rayleigh resolution limit [19] which states that two sources are resolvable when its spatial resolution satisfies:

$$R_{Rayleigh} \approx \frac{d}{D} \cdot \lambda \tag{4.10}$$

where d is the distance from the array to the source, D is the array aperture and λ is the wavelength of the acoustic signal source. This relationship assumes that the array aperture is significantly larger than the acoustic wavelength ($D >> \lambda$). However the Rayleigh resolution limit in Eq. (4.10) does not accurately predict the actual resolvability for our implementation. Instead simulations show that two sources become distinguishable only when the distance between sources are larger than one wavelength. This observation is consistent and summarized in Table 4.1, where a check-mark indicates successful separation of the two sources.



(a) Unresolved **(b)** Resolution Limit **(c)** Resolved

Figure 4.14: Beamformer output for two sources spaced 1 m apart, evaluated at (a) 250 Hz, (b) 315 Hz, and (c) 400 Hz respectively.

Hz \ m	0.15	0.2	0.25	0.5	1	3	5	7	9
40	✗	✗	✗	✗	✗	✗	✗	✗	✔
50	✗	✗	✗	✗	✗	✗	✗	✔	✔
63	✗	✗	✗	✗	✗	✗	✔	✔	✔
80	✗	✗	✗	✗	✗	✗	✔	✔	✔
100	✗	✗	✗	✗	✗	✗	✔	✔	✔
125	✗	✗	✗	✗	✗	✔	✔	✔	✔
160	✗	✗	✗	✗	✗	✔	✔	✔	✔
200	✗	✗	✗	✗	✗	✔	✔	✔	✔
250	✗	✗	✗	✗	✗	✔	✔	✔	✔
315	✗	✗	✗	✗	✔	✔	✔	✔	✔
400	✗	✗	✗	✗	✔	✔	✔	✔	✔
500	✗	✗	✗	✗	✔	✔	✔	✔	✔
630	✗	✗	✗	✔	✔	✔	✔	✔	✔
800	✗	✗	✗	✔	✔	✔	✔	✔	✔
1000	✗	✗	✗	✔	✔	✔	✔	✔	✔
1250	✗	✗	✔	✔	✔	✔	✔	✔	✔
1600	✗	✗	✔	✔	✔	✔	✔	✔	✔
2000	✗	✔	✔	✔	✔	✔	✔	✔	✔
2500	✔	✔	✔	✔	✔	✔	✔	✔	✔

Table 4.1: Spatial resolution.

4.5 Analysis

This section examine gathered technique-specific intermediate results, primarily from the validation metrics but also simulations and other valid inputs. The evaluation performed here highlights this techniques strengths and disadvantages and is the foundation for choosing a next suitable technique for implementation.

4.5.1 Impact of Microphone Placement Errors

The validation results presented in previous section reveal key insights into the performance characteristics and limitations of the DS beamformer under various offsets. The Monte Carlo simulations demonstrate that the DS beamformer exhibits high robustness to microphone placements, particularly at low and mid frequency ranges. In both on-axis and off-axis source configurations the DOA estimation remains reliable below 1600 Hz with the majority of estimates closely aligned with the true DOA. This is expected as lower frequencies correspond to longer wavelengths making the array configuration less sensitive to geometric deviations. As the frequency increases the beamformer becomes increasingly sensitive to positional inaccuracies.

4.5.2 Effect of Additive and Background Noise

In the presence of both additive Gaussian noise and background noise recordings the beamformer performance is degraded. Increased DOA estimation errors and degraded array responses are observed, especially at low and high frequencies. The presence of noise appears to reduce the presence of grating lobes. Likely due to the masking effect of increased noise levels. The frequency range of 250 - 1600 Hz demonstrates the highest robustness.

4.5.3 Sensitivity to Microphone Gain Variations

Simulations with varied microphone sensitives indicate minimal impact on DOA estimation accuracy compared to positional offsets. A gain offset of 20% shows no variation in the DOA estimation errors when the ULA is perfectly configuration. This is expected as the DS beamforming algorithm relies on alignment of the wavefront and not the relative amplitude of each channel.

4.5.4 Spatial Resolution Capabilities

The DS beamformer ability to resolve closely spaced sources is limited by both its aperture and operating frequency. The spatial resolution simulations show that the Rayleigh resolution limit underestimates the required distance for successful source separation in practice. Instead simulation results indicate the the sources must be separated by at least one wavelength to be distinguishable. This is clearly observed in Tabel 4.1 and Figure 4.14 where successful separation occurs above 315 Hz for a 1 meter source spacing.

4.5.5 Summary

Overall the DS beamformer performs well due to configuration imperfections but poorly in high noise environments, particularly in the low and high frequency range. However its resolution and high frequency sensitivity limitations should be acknowledged when interpreting results for high accuracy localization or source separation tasks.

To address the challenges caused by noisy conditions a beamforming approach capable of suppressing interference while maintain high directionality is required. One such methods is the Minimum Variance Distortionless Response (MVDR) beamformer which offers enhanced spatial filtering capabilities. This method is explored in detail in the following chapter.

5

Minimum Variance Distortionless Response Beamforming

This is the second technique chapter, it presents the Minimum Variance Distortionless Response (MVDR) beamformer. The chapter has the same structure as the previously evaluated DS beamformer, with the goal of evaluating the MVDRs performance. The strengths and weaknesses of the MVDR will be analysed and be the foundation for selecting the third technique to implement.

5.1 Technique Description

The decision to implement MVDR in this specific situation is based on the algorithms ability to handle noisy environments and its improved source separation capabilities compared to DS the beamformer. It is an adaptive beamforming technique, which means that it relies on the recorded microphone signals in order to calculate weights which are then applied to the microphone signals to create its output. MVDR is probably one of the most common adaptive beamformer algorithms, and is characterized by an array pattern with distinct sharp peaks.

This technique is often associated with Capon's work, *High-resolution frequency-wavenumber spectrum analysis* [20], and is frequently mentioned as *Capon's beamformer* or *Capon/MVDR* or similar variations. Some authors distinguish between Capon's method and MVDR citing them as being similar but different [17]. Across different research fields there also seems to be some naming inconsistency where this technique can be found under different names.

5.2 Theoretical Background

This subchapter presents the theory related to the MVDR beamforming technique and its application in this specific setup and scenario. The theoretical foundation is primarily based on the work and knowledge of Johnson and Dudgeon from their book *Array Signal Processing* [17] and Marc Lichtman's webpage `pysdr.org` specifically the section related to MVDR beamforming [21].

The MVDR beamforming technique applies weights $\boldsymbol{w}$ to recorded microphone signals $\boldsymbol{y}(f)$ to produce the beamformer output $z(f)$ for individual frequency bins as:

$$z(f) = \boldsymbol{w}^H \boldsymbol{y}(f) \tag{5.1}$$

In the previously implemented technique, DS beamforming, the weights applied per microphone signal are always $\boldsymbol{w} = 1$ and a steering vector $\boldsymbol{a}(\theta)$ is applied per microphone signals which steers the microphone arrays response towards a desired location.

In MVDR a Cross Spectral Matrix (CSM) is part of the implementation and it represents a similarity measurement between the individual microphone signals. The Combination of the CSM and steering vectors $\boldsymbol{a}(\theta)$ forms the basis for calculating the MVDR weights $\mathbf{w}$ which is applied to the microphone signals, as in equation 5.1. This process, which derives the weights based on information from the input signals via the CSM is what characterizes MVDR as an adaptive algorithm, compared to the conventional DS beamformer.

The MVDR beamformer can be formulated as a minimization problem with angle constraints related to a specific look direction, see equations 5.2. What the beamformer essentially does is setting the gain to unity in the look direction, $\boldsymbol{w^H a}(\theta) = 1$ and then striving to minimize the interference and noise coming from other directions.

$$\min_{w} \quad |\boldsymbol{w}^{\mathrm{H}}\boldsymbol{y}|^2 = \boldsymbol{w}^{\mathrm{H}}\boldsymbol{R}\boldsymbol{w} \quad \text{subject to} \quad \boldsymbol{w}^{\mathrm{H}}\boldsymbol{a}(\theta) = 1 \tag{5.2}$$

5.2.1 Cross Spectral Matrix

The CSM, denoted as $\boldsymbol{R}$ here, is a key component of adaptive beamforming technologies. It represents a similarity measurement between individual microphone signals recorded by a microphone array and their relationship is quantified and utilized in order to calculate weights for the beamformer. The matrix can be constructed in either time or frequency domain, in this specific context its constructed in frequency domain as:

$$\boldsymbol{R}(f) = \boldsymbol{y}(f)\boldsymbol{y}^{\mathrm{H}}(f) \tag{5.3}$$

Where $\boldsymbol{y}(f)$ represents the frequency domain representation of the input microphone signals arranged in a matrix where each row represents an individual microphone signal and each column value a frequency bin of signals. This ensures that $\boldsymbol{R}$ has the dimension of number of microphones x number of microphones, where each value represents a relationship between the individual microphone signals.

It is common to see beamforming literature referencing a covariance matrix or an estimated covariance matrix [22] [23] [24] [25], the implementation in equation 5.3 is similar to a normal estimated version. The $\boldsymbol{R}$ matrix in this situation is constructed similarly to how covariance is typically estimated, $\mathrm{cov}(A, B) = \sum(A - E(A))(B - E(B))$ but assuming that the mean values $E(A)$ and $E(B)$ is zero-valued simplifies the expression to $\sum AB$. In this setup, $A = \boldsymbol{y}(f)$ and $B = \boldsymbol{y}^{\mathrm{H}}(f)$ which produces the same expression as seen in equation 5.3. Due to implementation, the $\boldsymbol{R}$ matrix is calculated independently per frequency and never summed, therefore $\sum$ is left out.

There is some debate regarding the naming of this beamforming technique, whether *variance* is the correct term or if *power* would be more appropriate, thus if the beamforming algorithm should be named Minimum Variance Distortionless Response (MVDR) or Minimum Power

Distortionless Response (MPDR). This distinction depends on the signal content of which the $\boldsymbol{R}$ matrix is constructed. The signal content used to construct the $\boldsymbol{R}$ matrix is what settles the debate. Equation 5.4 shows how the the matrix could be decomposed into a source signal $\boldsymbol{R}_s$ component and a $\boldsymbol{R}_n$ noise signal component. The classification of the beamformer as either MVDR or MPDR depends on which of these components is used in the beamforming algorithm.

$$\boldsymbol{R}_z = \boldsymbol{R}_s + \boldsymbol{R}_n \tag{5.4}$$

The CSM of the MVDR beamformer is based on solely the noise component $\boldsymbol{R}_n$, whereas the MPDR beamformer is based on $\boldsymbol{R}_z$, containing both the source signal and the noise component [26]. In this specific implementation, with a stationary truck placed on a dynamometer, it is not practically feasible to separate source signals from noise components from data gathered during a pass-by measurement. As a result, our implementation aligns with the formulation of the CSM related to the MPDR beamformer, using $\boldsymbol{R}_z$. However since MVDR is a more commonly used technique and term and mixed definitions are common, we have chosen to continue referring this technique as MVDR.

An example of a normalized $\boldsymbol{R}$ matrix can be seen in figure 5.1. The signals are simulated with a sinusoidal source at 400 Hz located in front of microphone 11 with a small amount of noise added to the system. With knowledge of the microphone positions in the room the $\boldsymbol{R}$ matrix can by itself be used for source localization interpreting it similarly to a grid-based output from a beamformer.

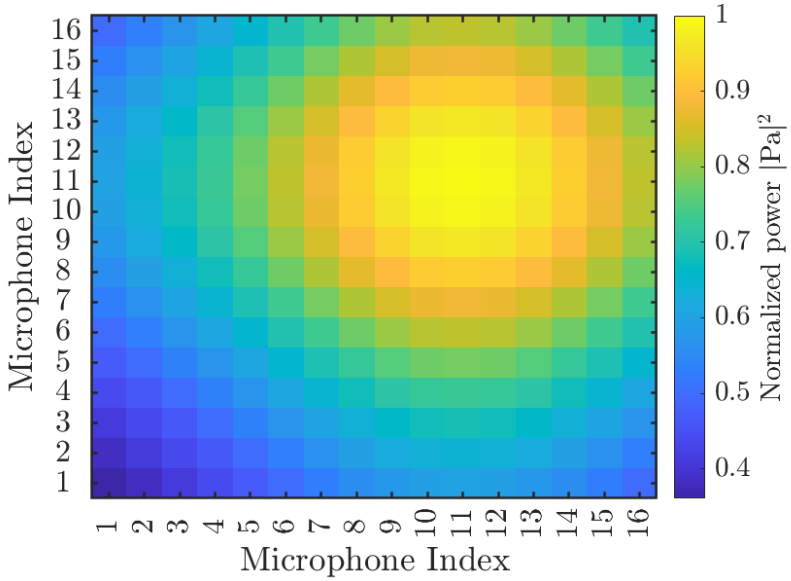


Figure 5.1: Visualization of CSM matrix as grid plot with normalized power on z-axis.

5.2.2 Adaptive Diagonal Loading

Calculating the MVDR beamformers weights the $\boldsymbol{R}$ matrix needs to be invertible. Unfortunately it is often ill-conditioned so that a inversion $\boldsymbol{R}^{-1}$ is not numerically stable. A

standard solution to this problem is to utilize Diagonal Loading (DL) which adds a small numerical value to every diagonal element of $\boldsymbol{R}$, which essentially is adding a small amount of noise to the signal in order to ensure the $\boldsymbol{R}$ matrix stability for inversion. The standard DL formulation can be expressed as:

$$\boldsymbol{R} = \boldsymbol{R} + \lambda \boldsymbol{I} \tag{5.5}$$

Where $\boldsymbol{R}$ is the CSM, λ is a loading factor - usually a small value, and the identity matrix $\boldsymbol{I}$. Changing λ affects how much noise is added to the diagonal which ensures invertibility at the same time affecting the beamformers array pattern resolution.

Alternative adaptive DL concepts were explored to simultaneously ensure the invertibility of $\boldsymbol{R}$ and maintain a precise array pattern. The goal was to implement a method which is repeatable and minimizes the amount of parameter tweaking due to varying input signal content. The approach selected and implemented is a simplified constraint-based method which imposes limits on the loading factor λ derived from statistical properties of the $\boldsymbol{R}$ matrix. A method proposed by authors Ma and Goh serves as a theoretical basis [22], their constraints imposed on the loading factor is stated as:

$$\sigma\left(\mathrm{diag}(\boldsymbol{R})\right) \leq \lambda < \overline{\mathrm{diag}(\boldsymbol{R})} \tag{5.6}$$

This equation indicates that λ should be larger then diagonal standard deviation σ and smaller then the diagonal mean of of $\boldsymbol{R}$. It is common to see this upper constraint expressed as $\frac{\mathrm{Tr}(\boldsymbol{R})}{N}$, where Tr denotes the trace function (sum of the diagonal elements) and N is the number of microphones.

The referenced article suggests an implementation of $\boldsymbol{R}$ where they average $\boldsymbol{R}$ over multiple snapshots, which builds a statistical basis over multiple samples. This approach is not applicable in our implementation, $\boldsymbol{R}$ is constructed per frequency and kept isolated as the weights $\boldsymbol{w}$ are calculated per frequency and applied as a filter in order to achieve a precise broadband MVDR implementation. This design has the implication that the statistical basis for constraining λ is limited due to the available content of just one $\boldsymbol{R}$ matrix per frequency. Due to the small sample size per frequency, the proposed constraints proposed in equation 5.6 often breaks down and its not uncommon for the lower bound of the constraint to exceed the value of the upper bound.

The simplest and most effective solution was to remove the lower bound and apply a tolerance factor to ensure that λ never exceeds the upper bound. Thus, the simplified constraint can be reformulated as as:

$$\lambda < \overline{\mathrm{diag}(\boldsymbol{R})} \tag{5.7}$$

This constraint can then be utilised to formulate an expression for calculating λ as:

$$\lambda = \epsilon \cdot \overline{\mathrm{diag}(\boldsymbol{R})} \tag{5.8}$$

This formulation requires $\epsilon < 1$ to satisfy the simplified constraint expressed in 5.7. This approach still involves one parameter that could be tuned ϵ, simulations and experiments has

proved that this implementation is quite robust. This is primarily due to the adaptive scaling provided by utilising the mean of the diagonal of $\boldsymbol{R}$, which acts as a relevant signal scaler.

A thorough review of scientific literature has not been able to identify any scientific sources that describes this exact adaptive DL setup. The only source found using the same formulation is a design user guide from Texas Instruments, where a Capon beamformer is designed and adaptive DL in the same format is applied [20].

5.2.3 Weights

The weights $\boldsymbol{W}$ is the central part of the MVDR algorithm and it is what the beamformer applies to the microphone signals in frequency domain $\boldsymbol{y}(f)$ in order to produce a beamformed output signal $\boldsymbol{z}(f)$ as seen in equation 5.1. The MVDR weights equation can be expressed as:

$$\boldsymbol{W} = \frac{\boldsymbol{R}^{-1}\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)\boldsymbol{R}^{-1}\boldsymbol{a}(\theta)} \tag{5.9}$$

The matrix $\boldsymbol{R}^{-1}$ refers to the inverse of previously described CSM. The steering vectors $\boldsymbol{a}(\theta)$ are formulated in the same fashion as for the DS beamformer, as described in equation 4.5 in subchapter 4.2. The DS beamformer used relative delay between individual microphones $\Delta\tau_{mk}$. MVDR steering vector time delay is instead formulated based on the direct source to individual microphone distance as it is expressed in equation 4.2b.

With all information gathered the weight $\boldsymbol{w}$ equation 5.9 can be updated to contain the correct CSM in regards to both signal content and applied adaptive DL, here expressed in matrix format as:

$$\boldsymbol{W} = \frac{(\boldsymbol{R}_z + \lambda\boldsymbol{I})^{-1}\boldsymbol{a}(\theta)}{\boldsymbol{a}^{\mathrm{H}}(\theta)(\boldsymbol{R}_z + \lambda\boldsymbol{I})^{-1}\boldsymbol{a}(\theta)} \tag{5.10}$$

This equation calculates the weights $\boldsymbol{W}$ for a certain frequency and for a set off scan angles θ. $(\boldsymbol{R}_z + \lambda\boldsymbol{I})^{-1}$ has size microphones x microphones and updates per frequency where λ is calculated as prescribed in equation 5.8. The steering vector $\boldsymbol{a}(\theta)$ has the shape microphones x 1 where it contains one value per individual microphone and angle. This solves for a weight vector $\boldsymbol{w}$ containing microphones x 1 values.

5.2.4 Output Power

Once the weights $\boldsymbol{W}$ is computed the output power of the MVDR beamformer for a given scan direction can be evaluated using the MVDR spectrum [27]. This represents the beamformer output power when steered toward a specific θ at a certain frequency and is formulated as:

$$\boldsymbol{p}_{\mathrm{MVDR}}(\boldsymbol{\theta}) = \frac{1}{\boldsymbol{a}^{\mathrm{H}}(\theta)(\boldsymbol{R}_z + \lambda\boldsymbol{I})^{-1}\boldsymbol{a}(\theta)} \tag{5.11}$$

This formulation corresponds to the output power of a spatial filter that enforces unit gain in the look direction while minimizing the contribution from other directions. Peaks in $\boldsymbol{p}_{\mathrm{MVDR}}(\boldsymbol{\theta})$ indicate the estimated DOA for incident wavefront at frequency f. The MVDR

spectrum is computed independently for each frequency and scan angle which can visualize frequency selective source localization and directional power estimation.

5.3 Implementation

The MVDR beamformer function which has been designed and implemented is a broadband frequency domain beamformer. Figure 5.2 shows a flowchart of the function.

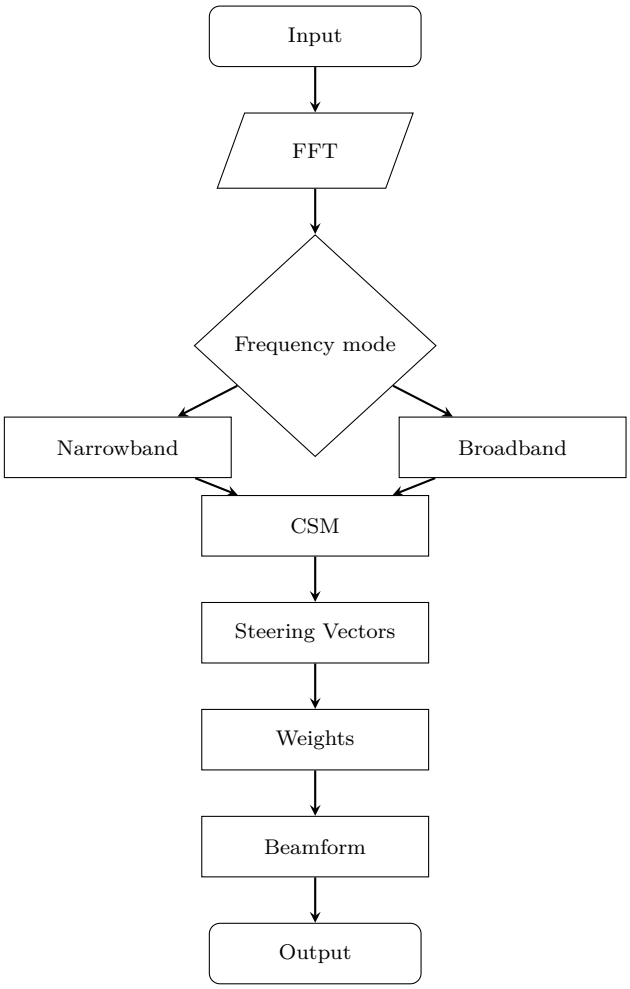


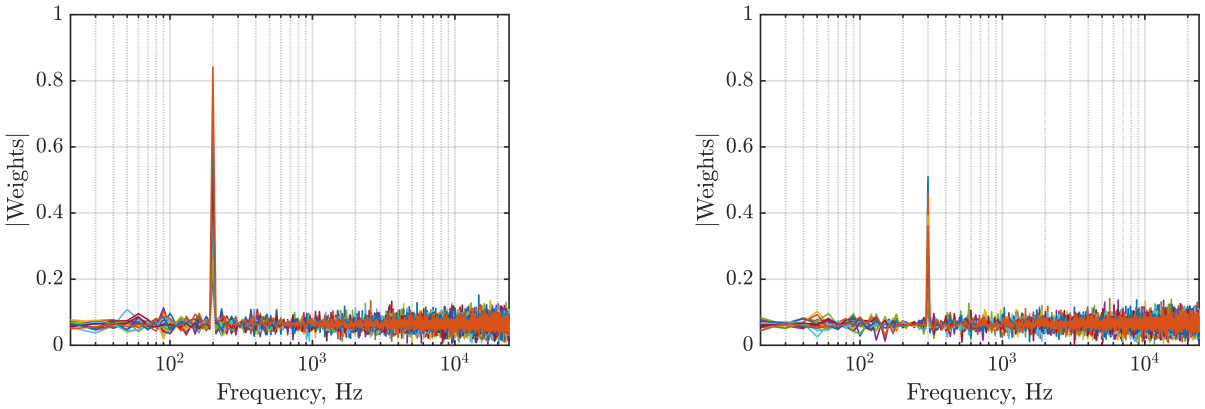
Figure 5.2: MVDR Beamformer Flowchart

The **Input** step takes the input signal, scan angles, sample rate and speed of sound together with spatial considerations as microphone placement and distance from scanning position. The **FFT** converts time domain input signals into frequency domain. **Frequency Mode** lets the user choose if the beamformer should operate in narrowband or broadband mode. Narrowband needs an extra frequency range input, effectively acting as a bandpass filter on the input signal. Broadband mode always operates in frequency range 0 Hz to half sampling rate $f_s/2$ Hz.

Next step calculates the **CSM** as described in subsection 5.2.1 with additional DL applied for invertibility. Near-field time delay based **Steering Vectors** is calculated for all microphones frequencies and angles.

All input data to calculate the MVDR beamformer **Weights** according to equation 5.10 is now gathered. The formula describes how the weights are calculated for one frequency and for one microphone. Practically the weights are stored in a 3D matrix format $\boldsymbol{W}(M, f, \theta)$ which contains individual weights values per microphones M for all calculated frequencies f and all scan angles θ. The 3D matrix will then have the shape microphones x frequencies x scan angles, see equation 5.12 which reformulates mentioned weights equation with all indices visible. Figure 5.3 shows an example of weights per individual microphones with two sources located at different positions in a simulation. Each plot relates to a specific steering angle.

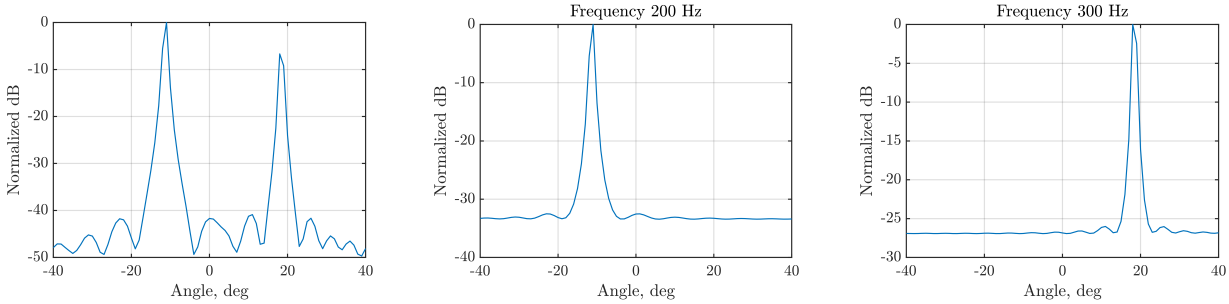
$$\boldsymbol{W}(M, f, \theta) = \frac{(\boldsymbol{R}_z(f) + \lambda \boldsymbol{I})^{-1} \boldsymbol{a}(M, \theta)}{\boldsymbol{a}^{\mathrm{H}}(M, \theta)(\boldsymbol{R}_z(f) + \lambda \boldsymbol{I})^{-1} \boldsymbol{a}(M, \theta)} \tag{5.12}$$



(a) Global bearing angle $-11°$ **(b)** Global bearing angle $18°$

Figure 5.3: Visualized MVDR weights for all microphones, frequencies in broadband frequency mode 0 to $f_s/2$ Hz. Two sources at different positions at frequencies $f = 200,\ 300$ Hz and $SNR = 10$ dB

In the **Beamform** step the calculated weights are applied to the input signal in frequency domain. In the **Output** stage both broadband and singular frequency SRP is calculated. In figure 5.4 individual frequency components related to a specific source frequency is plotted and the broadband summation of all calculated frequency components in range 0 to $f_s/2$ Hz. This example relates to the weights previously visualized in figure 5.3.



(a) Broadband SRP. **(b)** Individual frequency component 1. **(c)** Individual frequency component 2

Figure 5.4: SRP in broadband and per frequency narrowband representation for MVDR beamformer.

5.4 Validation Metrics

This section presents the validation results for the MVDR beamformer. Evaluation process follows the metrics defined in Chapter 3.4. The results presents MVDR beamforming behaviour across different test scenarios with focus on its ability to handle array element misplacements, sensitivity mismatches and to distinguish between closely spaced sources.

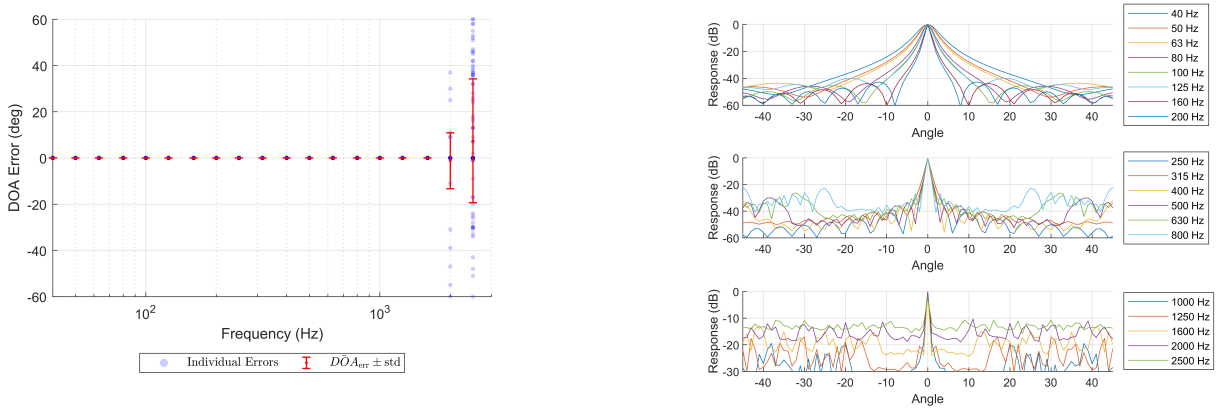
5.4.1 Robustness - Microphone Placements

The MVDR beamformer performance is evaluated under the same conditions as the DS beamformer. Using a 100 trial Monte Carlo simulation with microphone position offsets of up to ±5 cm in the x–y plane. The analysis is carried out under idealized and noisy conditions. Includes both on-axis ($\theta = 0°$) and off-axis ($\theta = -36°$) source positions. The results are used to assess the algorithm robustness to microphone misplacements by examining DOA estimation accuracy and array response consistency across frequency bands.

Ideal Case - on-axis

The MVDR beamformer DOA estimation error results under ideal on-axis conditions are presented in Figure 5.5a. Compared to the DS beamformer, the MVDR beamformer demonstrates similar results to microphone placement offsets across most frequencies in this idealized case. Below 2000 Hz the DOA estimation errors remain around the true source direction. The averaged array responses across all trials are shown in Figure 5.5. A clear difference here is that the mainlobe is much more narrower comparing with the DS beamformer responses. The side-lobes appear both lower in magnitude and more stable compared to DS results. This shows the MVDR beamformer ability to adaptively suppress contributions from non target directions even under positional uncertainty.

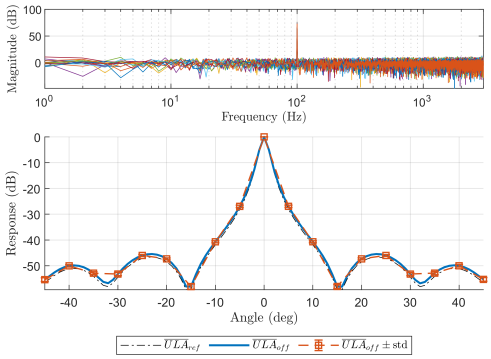
Frequency dependent variability in array responses is shown in Figure 5.6. At low frequencies the response remains nearly identical across trials, the same as DS beamformer. For higher frequencies such as 1000 Hz and 2000 Hz larger variability is introduced. However the mainlobe remains distinct and focused around its target.



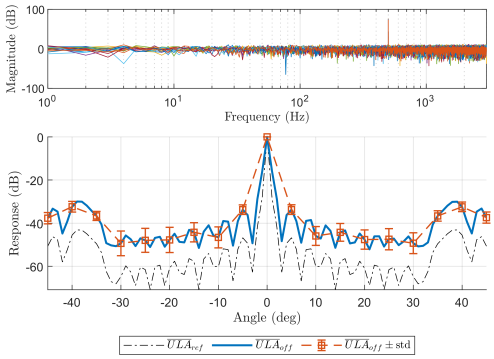
(a) Distribution of estimated DOA

(b) Average Array Responses

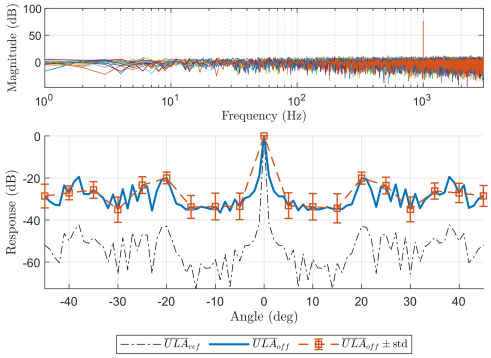
Figure 5.5: Performance of the MVDR beamformer in the ideal, noiseless scenario with the source placed on-axis. (a) shows the distribution of estimated directions of arrival (DOA) across all Monte Carlo trials, including the average DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.



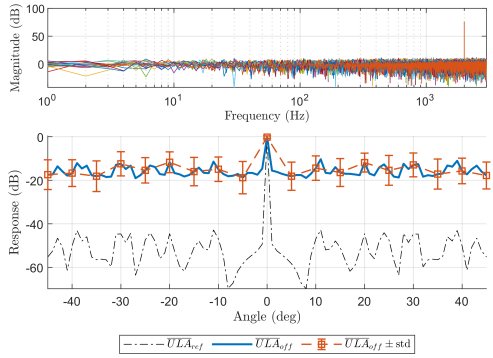
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz



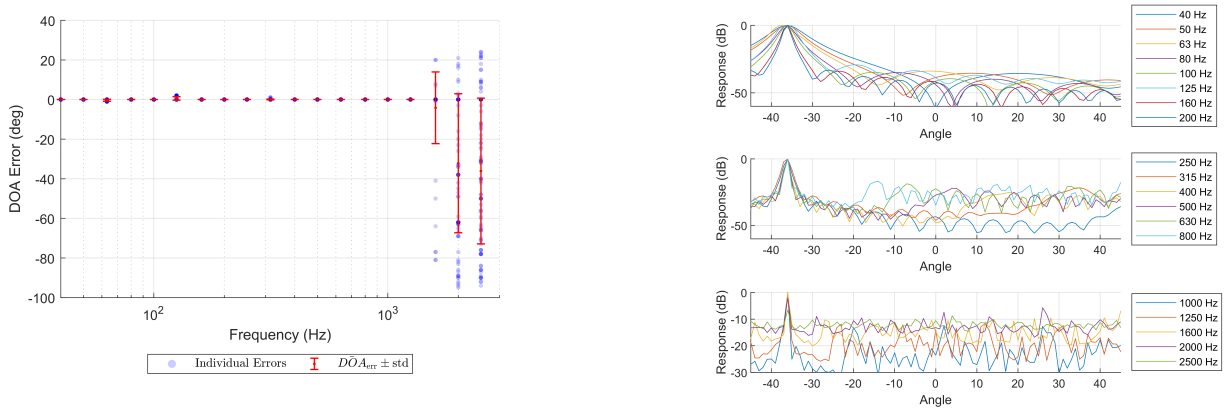
(d) 2000 Hz

Figure 5.6: MVDR beamformer array response variability in the ideal, noiseless scenario for an on-axis source at four frequencies.

Ideal Case - off-axis

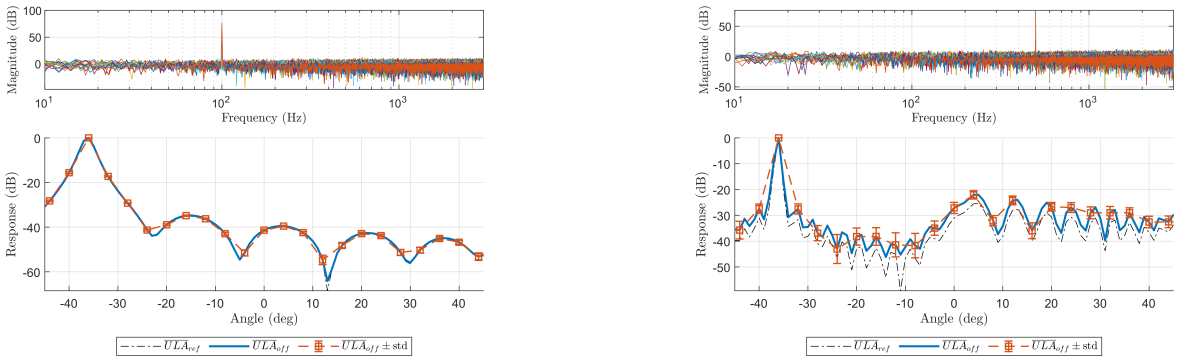
Evaluating the MVDR beamformes performance under more realistic conditions, the source was positioned off-axis at $\theta = -36°$. The DOA estimation error shown in Figure 5.7a demonstrate that the MVDR beamformer maintain accurate localization across most of the evaluated frequency bands. Up to around 1600 Hz the estimated DOA errors remain clustered around the true direction. At higher frequencies a moderate increase in spread is observed. The average array response remains well focused around the true DOA as seen in Figure 5.7b. Indicating moderate robustness to microphone placement offsets even in asymmetric source placements.

Figure 5.6 show the variability of the array responses across four representative frequency bands. The mainlobe gradually decrease in magnitude at higher frequencies, indicating increased sensitivity to misalignment and asymmetry. But the mainlobe still remains clearly visible.

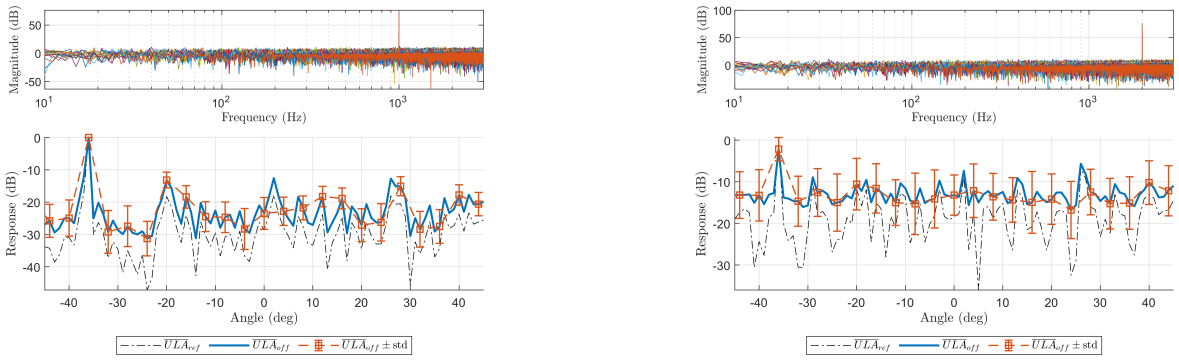


(a) Distribution of estimated DOA **(b)** Average Array Responses

Figure 5.7: Performance of the MVDR beamformer in the ideal, noiseless scenario with the source placed off-axis. (a) shows the distribution of estimated directions of arrival (DOA) across all Monte Carlo trials, including the average DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.



(a) 100 Hz **(b)** 500 Hz



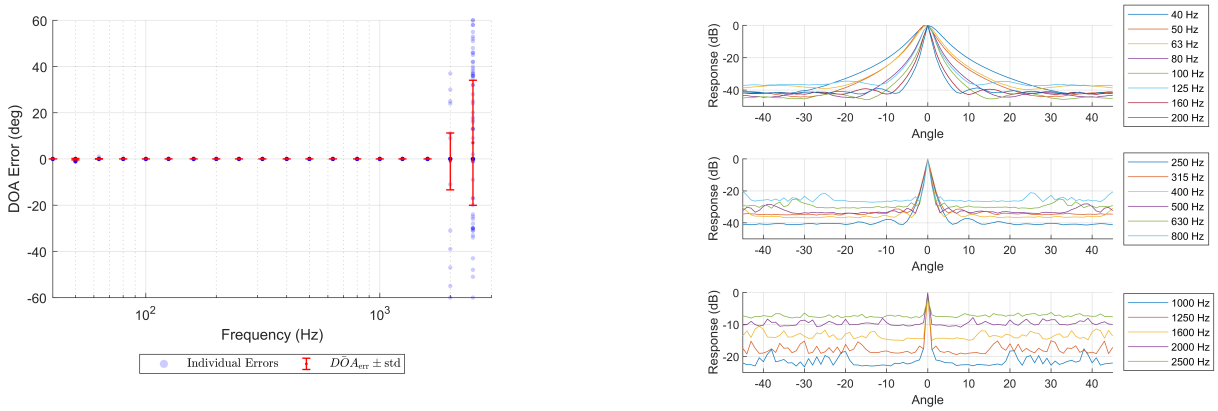
(c) 1000 Hz **(d)** 2000 Hz

Figure 5.8: MVDR beamformer array response variability in the ideal, noiseless scenario for an off-axis source at four frequencies.

Added Background & Additive Noise Case - on-axis

Under noisy on-axis conditions, the MVDR beamformer is assessed with both additive and background noise present in the system. The DOA estimation errors in Figure 5.9a show minimal change compared to the idealized condition. Unlike the DS beamformer which exhibited moderate estimation error spread under similar noise conditions, the MVDR beamformer effectively suppress noise contributions. This behaviour is clearly visible in the average array responses shown in Figure 5.9b where the side-lobes are strongly attenuated even with high noise present.This demonstrates the adaptive noise rejection capability of MVDR: The weights dynamically suppress energy from off-target directions preserving the mainlobe.

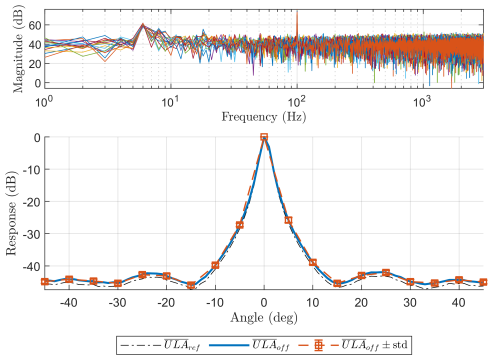
Figure 5.10 illustrates how microphone placement offsets reduce the SNR between the mainlobe and side-lobes. The comparison between the offset and reference arrays highlights the impact of positional accuracy, smaller placement errors yield improved directionality and clearer mainlobe separation.



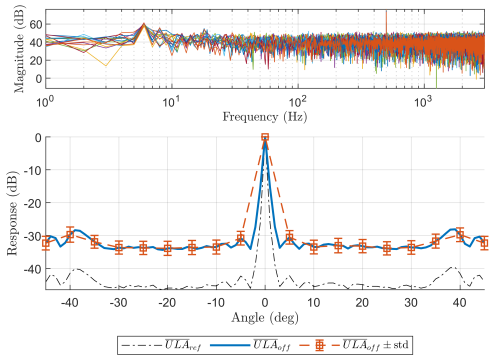
(a) Distribution of estimated DOA

(b) Average Array Responses

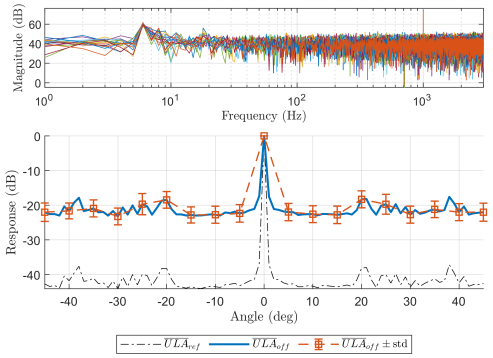
Figure 5.9: Performance of the MVDR beamformer with added noise for an on-axis source. (a) shows the distribution of estimated directions of arrival (DOA) across all Monte Carlo trials, including the average DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.



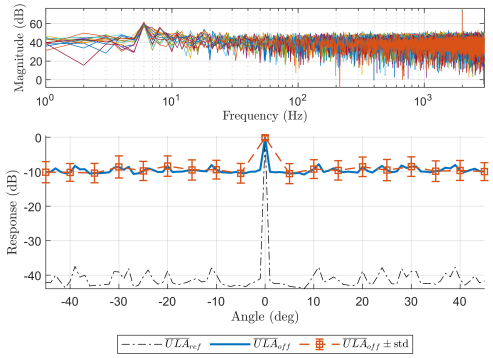
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz



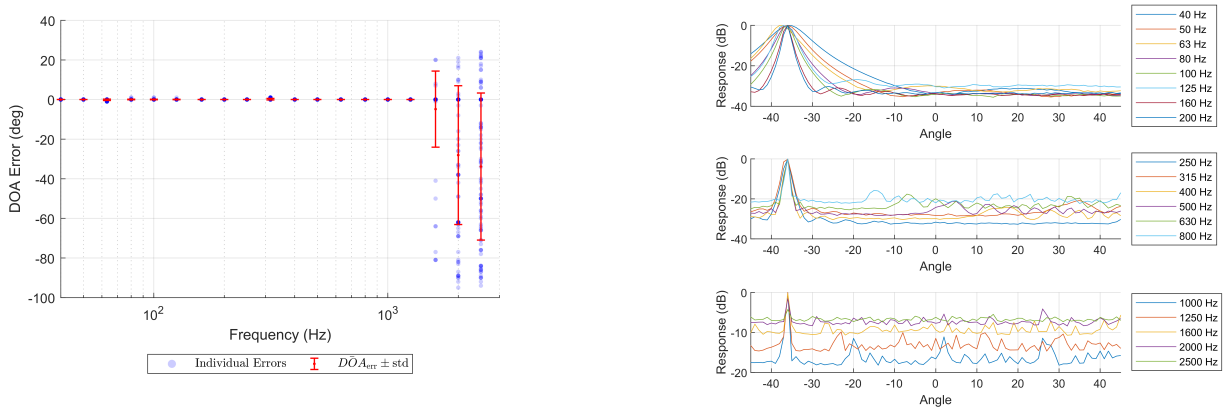
(d) 2000 Hz

Figure 5.10: MVDR beamformer array response variability with added noise for an on-axis source at four frequencies.

Added Background & Additive Noise Case - off-axis

With the source positioned off-axis at $\theta = -36°$ and background noise added. The MVDR beamformer continues to maintain reliable localization across most frequency bands. As shown in Figure 5.11a, DOA estimation errors remain concentrated near the true direction with deviations at higher frequencies. The average array response in Figure 5.11b confirms that the mainlobe remains correctly steered despite noise and microphone offsets.

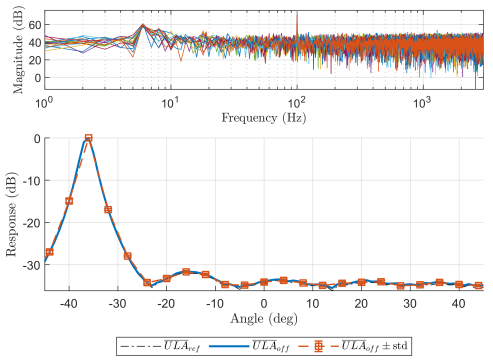
Figure 5.12 shows that variability increases slightly with frequency but the main beam remains focused around the target angle. This further confirms MVDR robustness under asymmetric and noisy conditions.



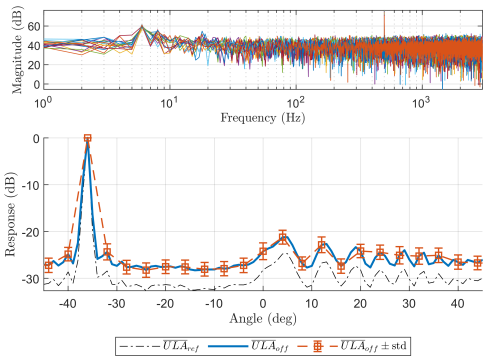
(a) Distribution of estimated DOA

(b) Average Array Responses

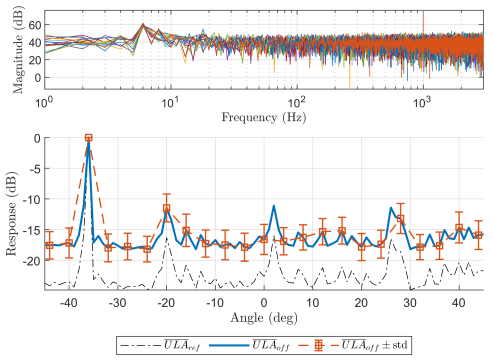
Figure 5.11: Performance of the MVDR beamformer with added noise for an off-axis source. (a) shows the distribution of estimated directions of arrival (DOA) across all Monte Carlo trials, including the average DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.



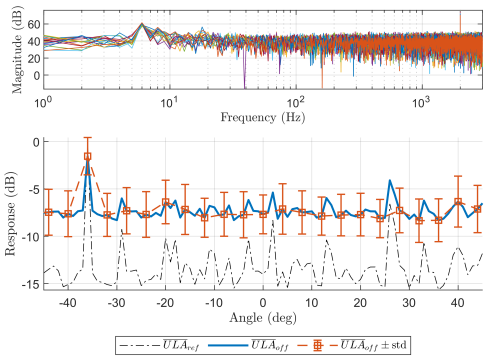
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz



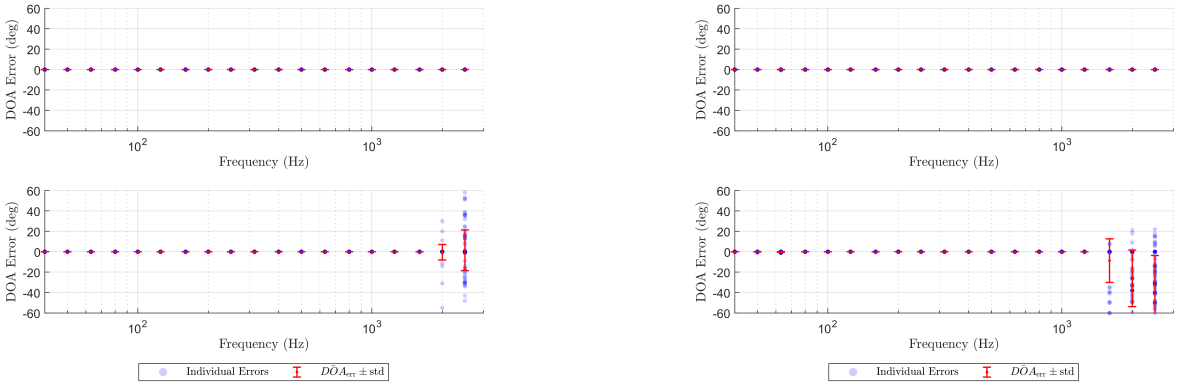
(d) 2000 Hz

Figure 5.12: MVDR beamformer array response variability with added noise for an off-axis source at four frequencies.

5.4.2 Robustness - Varied Microphone Sensitivity

This validation metric evaluates the MVDR beamformer sensitivity to microphone gain mismatches. The DOA estimation errors in Figure 5.13 show that for the reference ULA configuration the beamformer achieves perfect localization across all frequency bands for both on-axis and off-axis cases. These yield a zero estimation error, indication complete insensitivity to gain variation in the absence of geometric offsets of the ULA configuration.

For the offset ULA configuration estimation errors begin to appear around 1600 Hz and above. This spread is consistent with previous results where only positional offsets were introduced. Indicating that the dominant source of degradation is microphone placement errors, not sensitivity differences.

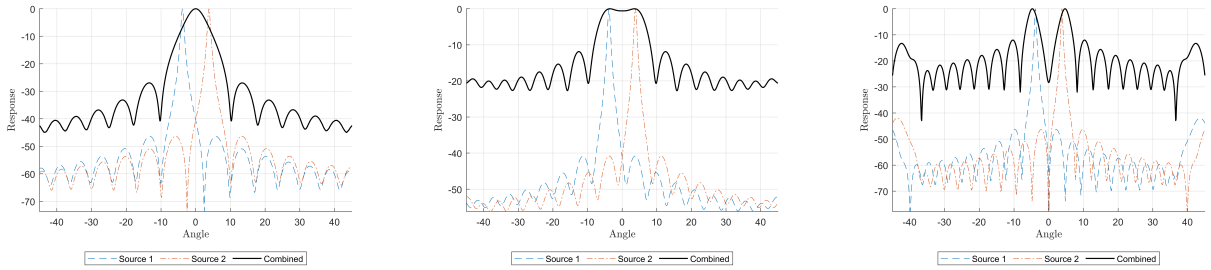


(a) On-Axis **(b)** Off-Axis

Figure 5.13: DOA estimation results under varied microphone sensitivity conditions for (a) on-axis and (b) off-axis source positions for MVDR beamformer.

5.4.3 Spatial Resolution

Altough the MVDR produces a narrower mainlobe compared to the DS beamformer its spatial resolution follows the same trend. As shown in Figure 5.14 the ability resolve two sources spaced 1 m apart improves with frequency. At 250 Hz the sources remain unresolved, at 350 Hz they begin to separate and by 400 Hz the two sources are clearly distinguished. The same resolution limit behaviour in the DS beamformer case applies here. This outcome is summarized in Table 4.1, which is valid for both beamformers. Despite improved directionality and noise suppression of MVDR the fundamental spatial resolution limit is still governed by source wavelength.



(a) Unresolved **(b)** Resolution Limit **(c)** Resolved

Figure 5.14: Beamformer output for two sources spaced 1 m apart, evaluated at (a) 250 Hz, (b) 315 Hz, and (c) 400 Hz respectively.

5.5 Analysis

The MVDR beamformer exhibits strong performance across all validation metrics. Outperforming the DS beamformer in the areas where it performance poorly, mainly under noisy conditions. Its core strength lies in its adaptive dynamic suppression of interference and noise while at the same time, preserving a distortionless response in the focus point direction. This is particularly advantageous in realistic environment where noise, microphone misplacements and gain mismatches are unavoidable.

5.5.1 Robustness to Microphone Placement Errors

Under ideal conditions with microphone position offsets MVDR maintains accurate DOA estimates for both on and off-axis placements up to approximately 2000 Hz. The beamformer adaptive suppression allows it to retain its mainlobe. Compared to DS the MVDR beamformer produces a sharper more stable array response at higher frequencies. At 1600 Hz and above degradation on performance becomes more apparent.

5.5.2 Performance in Noisy Conditions

A key advantage of the MVDR beamformer is its ability to maintain directivity in the presence of background and additive noise. Under noisy conditions the MVDR beamformer consistently preserves DOA estimation accuracy across the evaluated frequency bands. In contrast to the DS beamformer which exhibits increased estimation spread and degraded array responses MVDR effectively attenuates the noise and interference. This stems from the weighting process where the CSM allows the beamformer to distinguish between coherent and incoherent signal components. The successful implementation of adaptive diagonal loading ensures numerical stability.

5.5.3 Sensitivity to Microphone Gain Variations

Simulations involving varied microphone sensitives confirm that the MVDR beamformer is largely insensitive to magnitude mismatches. As long as the placement configuration of the array is preserved. When combined with geometric offsets performance trends mirror

the microphone placement validation. Showing that placement misalignments are dominant factor affecting MVDR accuracy. This support the fact that gain mismatches have negligible influence on MVDR output in practice.

5.5.4 Spatial Resolution Capabilities

Despite its narrower mainlobe width of the array response and improved noise and interference suppression the MVDR beamformer exhibit similar spatial resolution limits as the DS beamformer. This is evident in the resolution test where two sources spaced 1 m apart only become resolved above 350 Hz. The outcome confirms that while MVDR improves directionality and noise suppression the fundamental resolution limit is still governed by wavelength and array aperture.

5.5.5 Summary

The MVDR beamformer offer several advantages over conventional DS beamforming. Its robustness to noise and enhanced spatial selectivity make it well suited for complex and noisy environments. A notable strength of MVDR algorithm is the option to localize sources at specific frequencies. This enables the visualization of array responses for individual frequency bins as demonstrated in Figure 5.4b and 5.4c. In contrast the DS beamformer provides only a general broadband array response without frequency selectivity.

The benefits of MVDR comes at the cost of increased computational complexity and dependency on accurate estimation of the CSM. Despite these challenges, the validation results show that MVDR is a reliable and effective beamforming technique for applications requiring high localization accuracy and robustness to environmental uncertainty.

It is important to note that while MVDR improves directivity visualization and noise suppression its spatial resolution remains fundamentally limited by physical factors such as array aperture and signal wavelength. To further improve resolution additional methods are needed. To tackle this we explore the Deconvolution Approach for the Mapping of Acoustic Sources (DAMAS) which specifically targets resolution enhancement by deconvolving the array response. But before implementing DAMAS we first introduce an intermediate beamforming algorithm known as Functional Beamforming (FBF). FBF allows for adaptive control over array pattern through tunable exponent parameter and acts as a powerful preprocessing step that enhances the input to DAMAS.

6
Functional Beamforming

This is the third technique chapter and it presents the final beamformer part of this thesis, the Functional Beamforming (FBF) technique. Although this technique was not planned to be part of the thesis originally, during the implementation and evaluation of the fourth upcoming technique DAMAS, the potential in the combination of DAMAS and FBF became apparent. The chapters follows the same structure as the previous technique chapters.

6.1 Technique Description

The technique builds upon a specific implementation of the DS beamformer utilising a CSM. The FBF beamformer then has an additional extension, an exponential order factor which modifies both the eigenvalues of the CSM and the main beamforming expression. This extension allows FBF to significantly reduce side-lobe levels. FBF bridges the performance between our previously implemented techniques DS and MVDR and combines their individual strengths in terms of both robustness and resolution. In 2014, Robert P. Dougherty introduced the FBF algorithm at the Berlin Beamforming Conference [28].

This technique was originally not thought to be a part of this master's thesis but during the work with the fourth upcoming technique DAMAS, FBF was examined as a possible input beamformer which could solve certain problems and improve DAMAS result, but what these problems are, is left as a cliffhanger for chapter 7. During the work with DAMAS we realized that FBF had potential as an additional beamformer technique because of its ability to reduce side-lobes while keeping source responses intact. Therefore it is a natural extension to, and has positive synergy effects with DS and MVDR, at the same times it offers tweakability of array patterns through the order factor which is a helpful addition.

6.2 Theoretical Background

The FBF technique is an extension of a specific DS beamformer implementation utilising a CSM matrix. Therefore this theoretical background chapter will first introduce the CSM based DS technique abbreviated as DS-CSM. Then the chapter will extend the DS-CSM beamformer into the FBF.

6.2.1 Delay-and-Sum Based on CSM

This section is a continuation of the previously implemented technique DS, see chapter 4. Here the algorithm is expressed fully in frequency domain representation and the goal is to explain a certain type of implementation which uses a CSM matrix of a signal instead of the actual time-domain signal as its input.

Building on the theory explain in the DS chapter, which is based on Johnsons and Dudgeon's DS beamformer [17]. The beamformer output in time domain can be expressed as seen in equation 4.1. This equation can be re-expressed in the frequency domain. Now with steering vector $a(\theta)$ lining up the individual microphone signals via phase-shifts instead of time-domain sample shifting. The frequency domain DS beamformer can be expressed for a certain angle θ and frequency as:

$$Z(\theta, f) = \sum_{m=0}^{M-1} wY(f)a(\theta) \tag{6.1}$$

where w are the beamformer weights, $Y(f)$ are the individual microphone signals in frequency domain, which is then summed over all elements to achieve a non-normalized output $Z(\theta, f)$.

The steering vectors $a(\theta)$ can be expressed as in equation 4.5 with direct delay τ_{m_k} as in equation 4.2b. There is an additional amplitude normalization term added, based on r_{m_k}, which is the distance related to τ_{m_k}, the distance between a beamforming focus position and a specific microphone. Also the central position of the array to the focus position r_c. The steering vectors from 4.5 can now be updated as:

$$a(\theta) = \frac{r_{m_k}}{r_c} e^{-j\omega\tau_{m_k}} \tag{6.2}$$

The frequency domain beamformer output can be reinterpreted in a matrix format and slightly adapted compared to Johnson & Dudgeon as:

$$Z(\theta, f) = \boldsymbol{a}^{\mathbf{H}}(\theta)\boldsymbol{W}\boldsymbol{Y} \tag{6.3}$$

Where $\boldsymbol{W}$ and $\boldsymbol{Y}$ has sizes microphones x microphones respectively microphones x frequency components of a signal represented in frequency domain. This can be used to express the SRP for a specific frequency as:

$$\boldsymbol{p}_{\text{DS-CSM}} = \boldsymbol{a}^{\mathrm{H}}\boldsymbol{W}\boldsymbol{Y}\boldsymbol{Y}^{\mathbf{H}}\boldsymbol{W}^{\mathbf{H}}\boldsymbol{a} \tag{6.4}$$

The DS beamformer is a non-weighted beamformer, where the weights are often expressed as $w = 1$ for every microphone in the array. However, this doesn't take into account normalization which is applied to ensure the beamformers output is properly scaled to compensate for the number of microphone signals recorded by the array. This normalization can be viewed as parts of the weights which then updates the DS beamformer weights expression per microphone as $w = 1/M$ where M is the total number of microphones, represented as a scalar value. This can be used to update the SRP equation 6.4 as:

$$\boldsymbol{p}_{\text{DS-CSM}} = \frac{\boldsymbol{a}^{\mathrm{H}}\boldsymbol{Y}\boldsymbol{Y}^{\mathrm{H}}\boldsymbol{a}}{M^2} \tag{6.5}$$

The term $\boldsymbol{YY^H}$ corresponds to the CSM which is introduced and denoted as $\boldsymbol{R}$ in the MVDR subchapter 5.2.1, equation 5.3. Additional DL techniques are added and applied to $\boldsymbol{R}$ to ensure invertibility. Here, there is no need for inversion, it would just add unnecessary noise to the CSM. Therefore, in order to avoid confusion, the CSM is renamed here as $\boldsymbol{C} = \boldsymbol{YY^H}$, it is the same type CSM but without any DL added. With the introduction of $\boldsymbol{C}$ the SRP of the DS-CSM beamformer can now be expressed as:

$$\boldsymbol{p}_{\text{DS-CSM}} = \frac{\boldsymbol{a}^{\text{H}}\boldsymbol{C}\boldsymbol{a}}{M^2} \tag{6.6}$$

Which almost coincides with the type of DS implementation utilized in an article by Wu and Chang [29] which is the input basis for the DAMAS algorithm in upcoming chapter 7.

6.2.2 Functional Beamforming

The functional beamformer builds upon the previously introduced DS-CSM beamformer and extends this model. This algorithm utilises an exponential order factor υ which can be used to influence and tune the array pattern of the FBF. It is added to the DS-CSM beamformer through eigenvalue decomposition of the CSM matrix $\boldsymbol{C}$ and applied on to the eigenvalues as an exponent in the beamformer expression.

The CSM matrix $\boldsymbol{C}$ is eigenvalue decomposed (EVD) into underlying matrices $\boldsymbol{P}$ and $\boldsymbol{D}$ which contains eigenvectors and eigenvalues respectively and can be expressed as:

$$\boldsymbol{C} = \boldsymbol{PDP^{-1}} \tag{6.7}$$

The eigenvalue matrix $\boldsymbol{D}$ is a diagonal matrix. The order factor is added as an exponent in the form $1/\upsilon$ per eigenvalue which attenuates larger eigenvalues compared to smaller ones and can be expressed as:

$$\boldsymbol{C}_\upsilon = \boldsymbol{P}\boldsymbol{D}^{\frac{1}{\upsilon}}\boldsymbol{P}^{-1} \tag{6.8}$$

The DS-CSM beamformer expression can now be expanded with this eigenvalue attenuated CSM version and with υ as an exponent to the overall beamformer expression. The SRP expression from equation 6.6 can now be updated:

$$\boldsymbol{p}_{\text{FBF}} = \frac{\left(\boldsymbol{a}^{\text{H}}\boldsymbol{C}_\upsilon\boldsymbol{a}\right)^\upsilon}{M^2} \tag{6.9}$$

This is the FBF, suggested by Dougherty [28] with an additional normalization factor based on the number of microphones part of the array M.

Order Factor

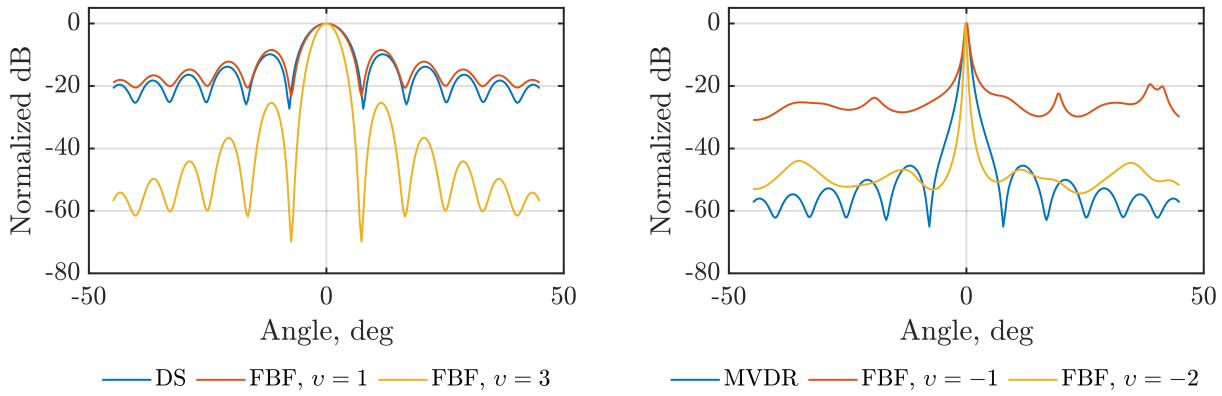
The order factor υ has some interesting properties which relates this method to both DS and MVDR. Dougherty says that when $\upsilon = 1$, the FBF has the same behaviour as the DS-CSM, when $\upsilon = -1$ the behaviour instead mimics an MVDR beamformer [28]. This behaviour is present in our implementation as well but it could be more thoroughly defined.

Positive υ mimics the rounded mainlobe behaviour which is present in our DS implementation, see figure 6.1a. Notice that the DS beamformer here, is the filter-based DS beamformer

described in chapter 4. This explains the small level differences exhibited comparing filter DS and FBF with $\upsilon = 1$.

Negative υ produce sharper and narrower MVDR-like lobes. Figure 6.1b shows MVDR and FBF with $\upsilon = -1, -2$. The $\upsilon = -1$ case exhibits MVDR-like sharp mainlobe, but it is first at $\upsilon = -2$, in this specific setup, that FBF experience similar side-lobe levels as MVDR.

Increasing $\upsilon \to \infty$ for DS-like and decreasing $\upsilon \to -\infty$ for MVDR-like behaviour continually sharpens the SRP, suppressing side-lobes and narrowing mainlobe width up until a certain υ where the beamformer breaks.



(a) DS and FBF.

(b) MVDR and FBF.

Figure 6.1: Behaviour comparison between implemented beamformers DS, MVDR and FBF with different adapted υ. Ideal simulation, Single source $f = 200$ Hz no noise added.

6.3 Implementation

The FBF beamformer algorithm is presented as containing two separate modules. The first is the DS beamformer implementation based on a CSM, previously referred to as DS-CSM. The second module is the FBF extension where EVD is performed and the order factor υ is applied to shape the beamformer output. Figure 6.2 contains a flowchart where both modules are highlighted by dashed rectangles.

The **Input** stage to the beamformer algorithm uses time domain microphone signals, scanning distance and angles, order factor, sample rate, speed of sound. Also spatial measurements as distance between central position of the array and a specific beamforming focus position r_c and from the same focus position but to every individual microphone r_{m_k}.

The input signal is converted to frequency domain in the **FFT** step and **Steering Vectors** are calculated as expressed in equation 6.2. The last step contained in the DS-CSM module is the calculation of the CSM matrix C based on the frequency domain input signals.

The FBF extension modules first step, the **EVD Order Factor** is the eigenvalue decomposition of C with added exponent v^{-1} applied to the eigenvalues as described in equation 6.8. The **Beamform** step calculates SRP as in equation 6.9 with v exponent addition. The **Output** step calculate both broadband and narrowband per frequency output where the broadband output contains all evaluated frequencies and narrowband output individual frequencies.

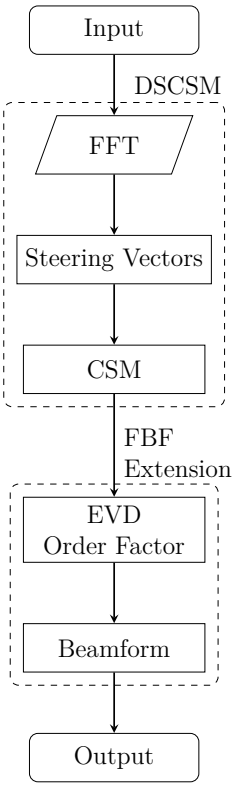


Figure 6.2: FBF Beamformer Flowchart

6.4 Validation Metrics

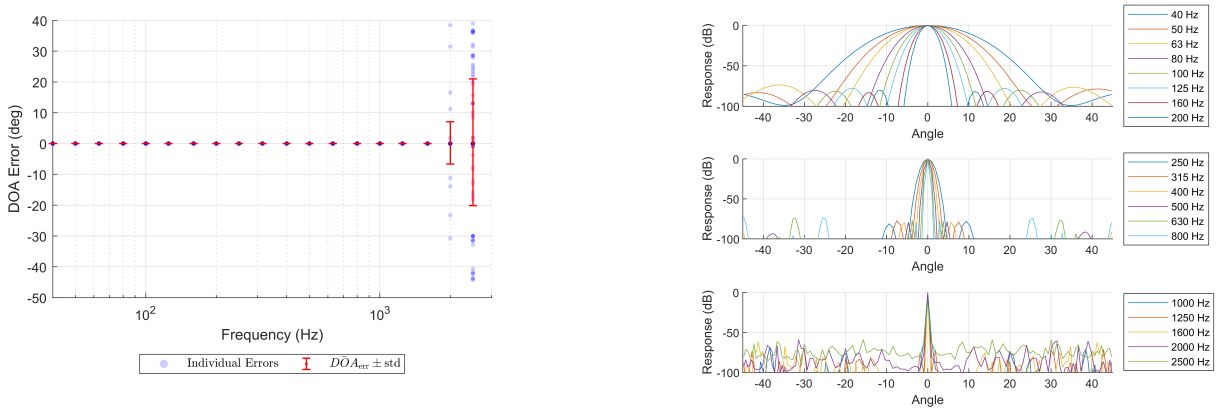
Validation of the FBF beamformer follows the same metric framework as previous configurations. The results show its performance across the evaluation scenarios. For the FBF simulations the order factor is set to 10. This was chosen to balance both performance and numerical stability. Allowing effective noise and artifacts suppression without degrading the beamformer resolution and directional sensitivity. At the same time its result is consistently separated from the DS and MVDR beamformers behavior.

6.4.1 Robustness - Microphone Placements

In this section robustness to microphone placement errors is presented and evaluated by introducing random offsets to each microphone position same way as previous beamforming techniques. This determines the FBF beamformers sensitivity to deviations from the ideal array geometry.

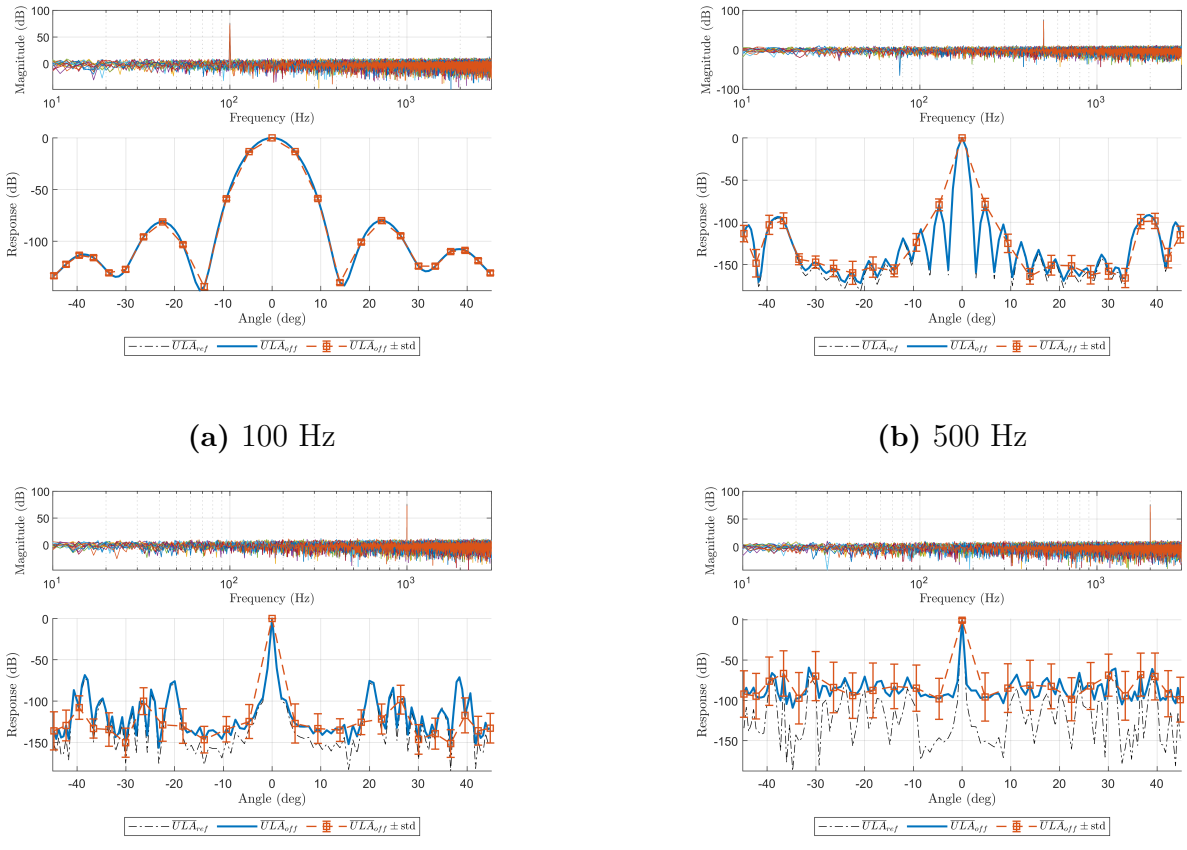
Ideal Case - on-axis

Figure 6.3 shows that the FBF beamformer maintain low DOA estimate errors across all frequency bands up to 2000 Hz. This estimation spread remains comparable to the DS and MVDR beamformers. The average array response is centered on the true direction with a clearly defined mainlobe and really low side-lobe levels. The performance shown in Figure 6.4 confirms high consistency across trials with some increased variability appearing above 1000 Hz.



(a) Distribution of estimated DOA **(b)** Average Array Responses

Figure 6.3: Performance of the FBF beamformer in the ideal, noiseless scenario with the source placed on-axis. (a) shows the distribution of estimated directions of arrival (DOA) across all Monte Carlo trials, including the average DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.



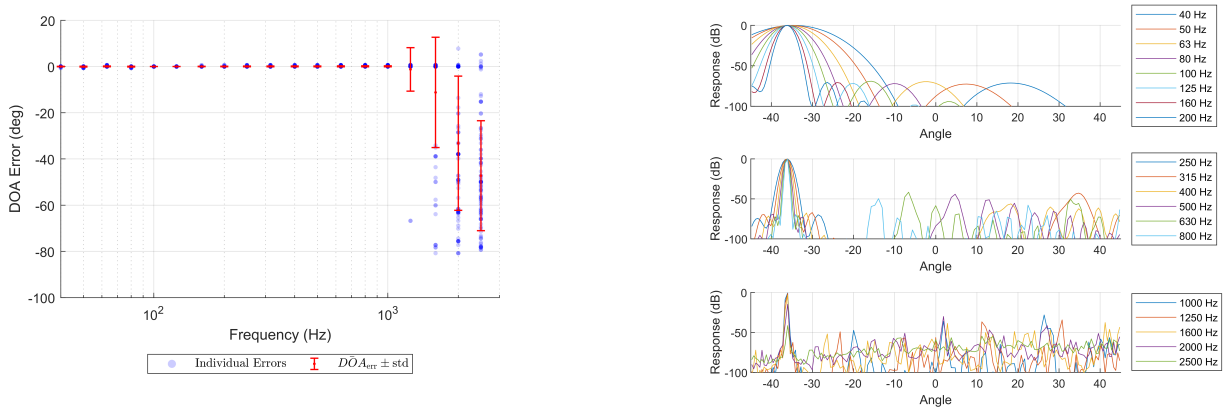
(a) 100 Hz **(b)** 500 Hz

(c) 1000 Hz **(d)** 2000 Hz

Figure 6.4: FBF beamformer array response variability in the ideal, noiseless scenario for an on-axis source at four frequencies.

Ideal Case - off-axis

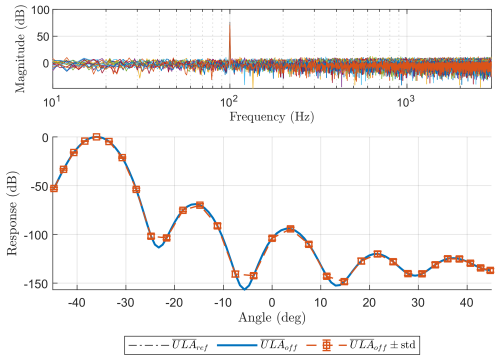
The DOA estimation results for the off-axis source ($\theta = -36°$) are shown in Figure 6.5a. The FBF beamformer retains an overall accurate localization across the frequency range up to 1600 Hz. The average array response in Figure 6.5b confirms that the beamformer correctly focuses on the off-axis direction with high SNR levels because of the high v. Increased variability is observed in the higher frequency bands in Figure 6.6 but the mainlobe remains distinguishable.



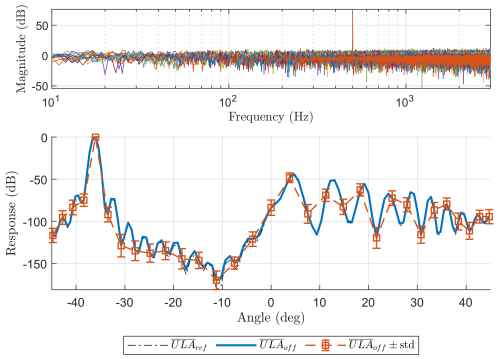
(a) Distribution of estimated DOA

(b) Average Array Responses

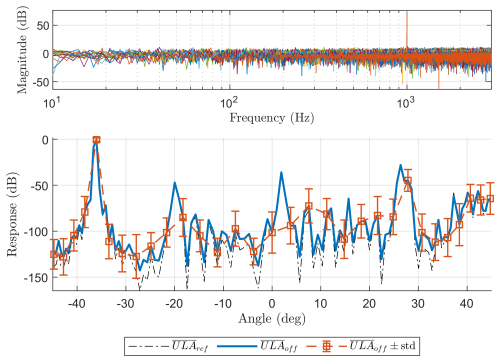
Figure 6.5: Performance of the FBF beamformer in the ideal, noiseless scenario with the source placed off-axis. (a) shows the distribution of estimated directions of arrival (DOA) across all Monte Carlo trials, including the average DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.



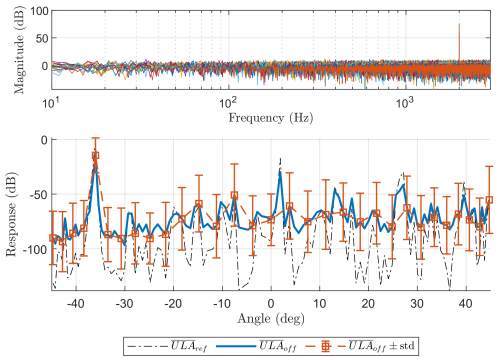
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz

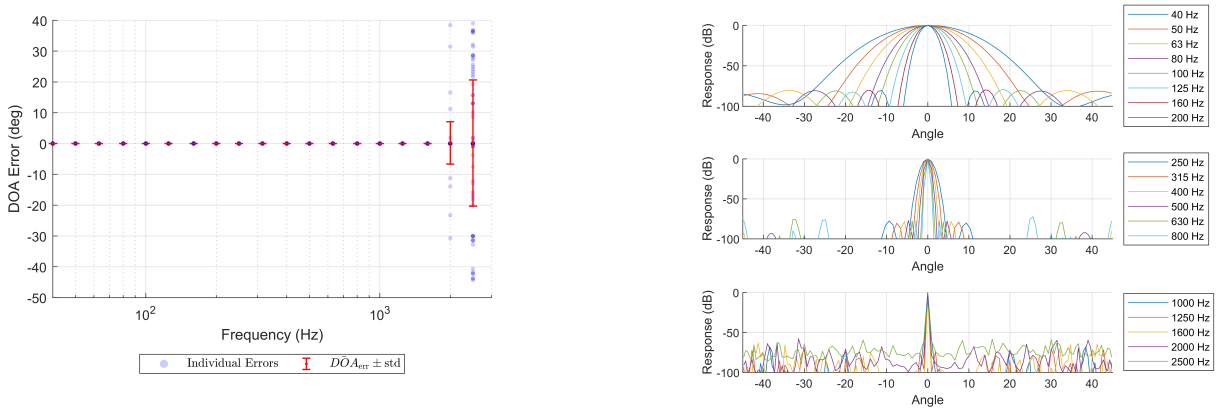


(d) 2000 Hz

Figure 6.6: Performance of the FBF beamformer in the ideal, noiseless scenario with the source placed off-axis. (a) shows the distribution of estimated directions of arrival (DOA) across all Monte Carlo trials, including the average DOA and its standard deviation. (b) presents the corresponding average array response patterns across all source frequencies.

Added Background & Additive Noise Case - on-axis

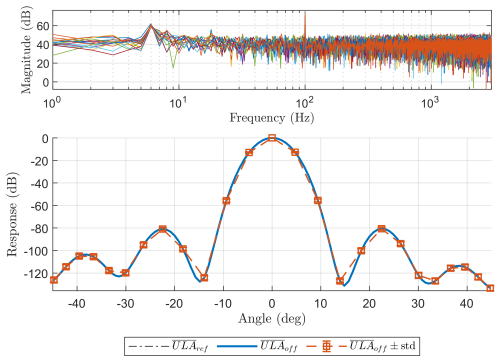
Under noisy conditions the FBF beamformer continues to localize the on-axis source effectively up to 2000 Hz as shown in Figure 6.7a. Figure 6.7b display that the mainlobe remains around the true DOA while side-lobes are suppressed. In Figure 6.8 increased noise result in spread at higher frequencies as expected but with high SNR levels even at 2000 Hz.



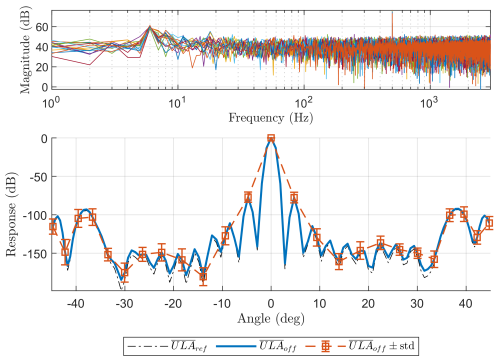
(a) Distribution of estimated DOA

(b) Average Array Responses

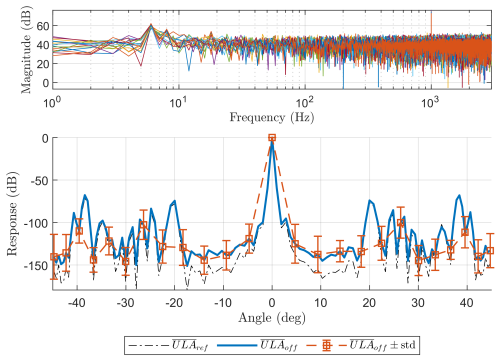
Figure 6.7: Performance of the FBF beamformer in the presence of additive and background noise when the source is placed off-axis. (a) illustrates the distribution of DOA estimates across all Monte Carlo trials, including the average estimated DOA and its standard deviation. (b) presents the corresponding average array response.



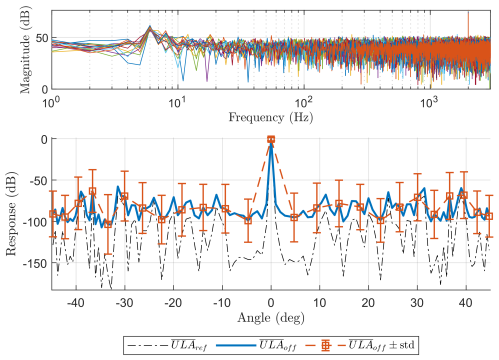
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz

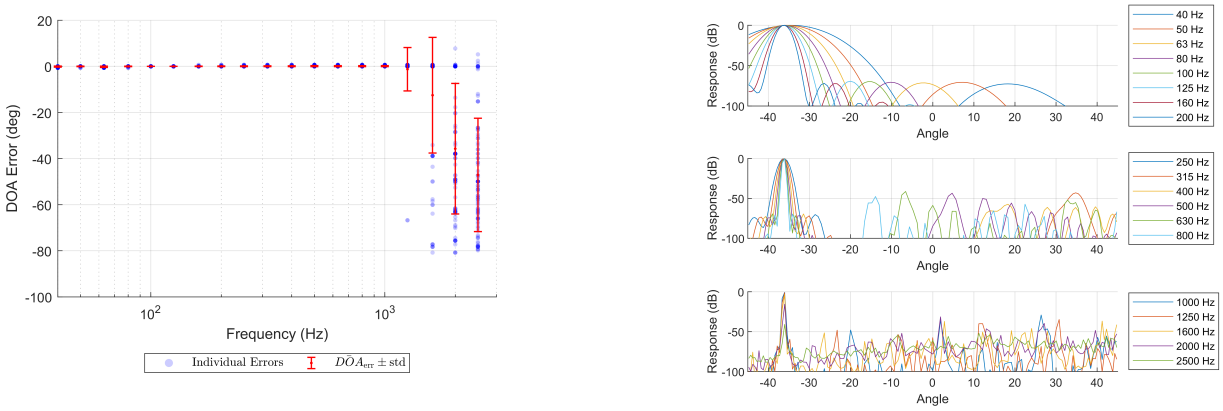


(d) 2000 Hz

Figure 6.8: FBF beamformer array response variability in the ideal, noiseless scenario for an off-axis source at four frequencies.

Added Background & Additive Noise Case - off-axis

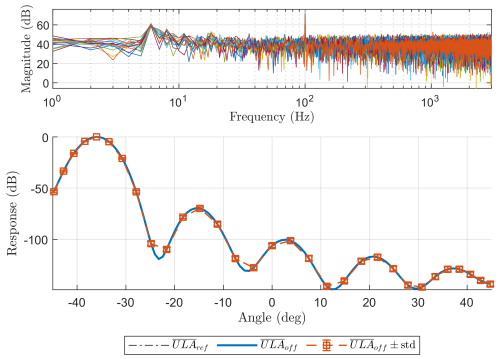
When the source is off-axis and high noise levels is added the FBF beamformer maintains high localization performance as seen in Figure 6.9a up to around 1600 Hz, at 1250 Hz, only one large DOA error appears. The corresponding array responses in Figure 6.10 confirm that the main lobe remains focused round its true source direction with slightly higher variability compared to the ideal case.



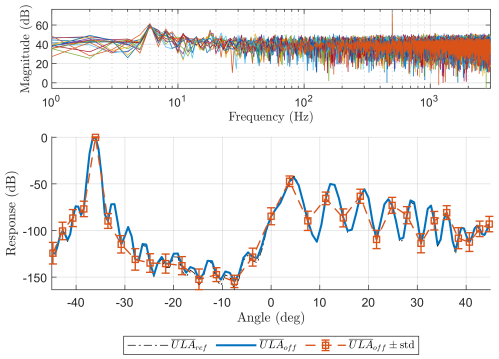
(a) Distribution of estimated DOA

(b) Average Array Responses

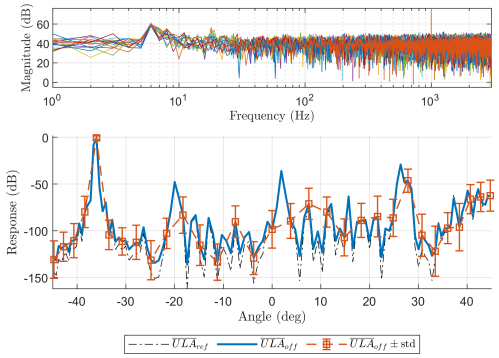
Figure 6.9: Performance of the FBF beamformer in the presence of additive and background noise when the source is placed off-axis. (a) illustrates the distribution of DOA estimates across all trials, including the average estimated DOA and its standard deviation. (b) presents the corresponding average array response.



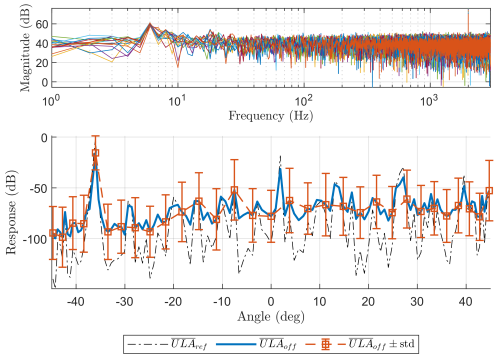
(a) 100 Hz



(b) 500 Hz



(c) 1000 Hz

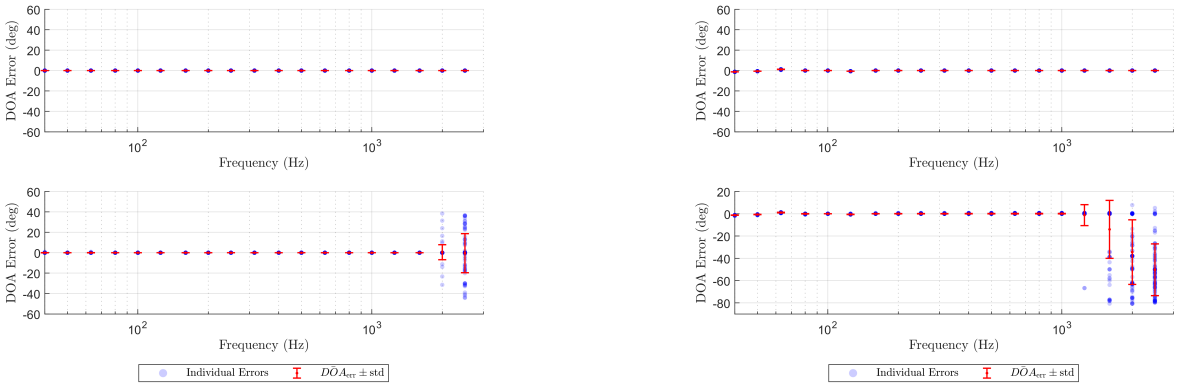


(d) 2000 Hz

Figure 6.10: FBF beamformer array response variability with added noise for an off-axis source at four frequencies.

6.4.2 Robustness - Varied Microphone Sensitivity

Under varied microphone sensitivity the FBF beamformer shows no degradation in localization accuracy for reference ULA configuration. As seen in top row of figure Figure 6.11a and Figure 6.11b. All DOA errors show no variation, indicating perfect source localization.
In the presence of placement offsets spread increase with higher frequencies, showing higher spread around 1600 Hz and upwards. These results are consistent with previously observed geometry errors. Sensitivity variations alone has negligible impact on beamforming performance.

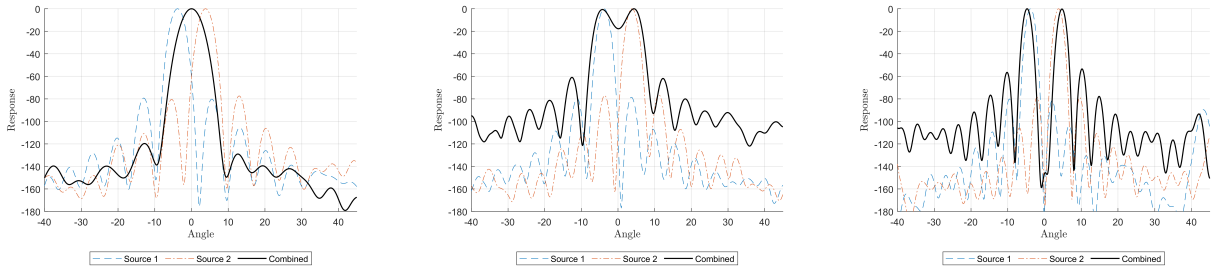


(a) On-Axis **(b)** Off-Axis

Figure 6.11: DOA estimation distribution for (a) on-axis and (b) off-axis source positions. The top row shows results using the reference ULA configuration. Bottom row presents results with microphone position offsets applied.

6.4.3 Spatial Resolution

The spatial resolution behaviour of the FBF beamformer is identical to that of the DS beamformer. The FBF exhibits the same resolution limits with respect to source separation. Figure 6.12 illustrates the transition from unresolved to resolved source pairs as frequency increases. As with the DS and MVDR beamformers. The results follow the pattern summarized in Table 4.1.



(a) Unresolved **(b)** Resolution Limit **(c)** Resolved

Figure 6.12: Beamformer output for two sources spaced 1 m apart, evaluated at (a) 250 Hz, (b) 315 Hz, and (c) 400 Hz respectively.

6.5 Analysis

The FBF technique offers a balance between robustness and spatial resolution by extending the DS beamformer into an eigenvalue structured solution. Its performance across all validation conditions demonstrate that it effectively combines key strengths of both DS and MVDR while introducing a tunable degree of control via the order factor υ.

6.5.1 Directional Performance

Under microphone placement errors FBF maintains accurate DOA estimation results across all frequencies up to 1600 Hz for both on and off-axis source configurations. The average array responses show a consistent narrow and well defined mainlobe with significantly lower side-lobe and grating lobe levels compared with DS and MVDR.

In the presence of additive and background noise the FBF continues to show strong directionality. Its DOA estimation accuracy and SNR between mainlobe and side-lobes remains very high. This suggests that FBF is well suited for noisy environments.

6.5.2 Sensitivity to Microphone Gain

FBF is insensitive to microphone sensitivity variations. In cases with gain mismatch but no positional error it achieves perfect localization across all frequencies, similar to DS and MVDR. When positional offsets are introduced alongside gain mismatches performance remains consistent with earlier tests. This confirms that the dominant degradation factor is array geometry and that FBF handles gain inconsistency sufficiently.

6.5.3 Spatial Resolution Capability

Despite improved mainlobe sharpness FBF does not surpass the fundamental resolution limits caused by array aperture and signal wavelength. Unlike DS and MVDR, FBF significantly increases the SNR between mainlobe and surrounding side-lobes. Although the resolution threshold itself is not fundamentally improved the tunable beamformer output provides a more structured and interpretable array response. This is beneficial for post-processing methods such as DAMAS where suppressing side-lobe levels improves the effectiveness of deconvolution and leads to more accurate source maps.

6.5.4 Summary

FBF serves as a extension to both DS and MVDR, providing more flexible design with tunable array pattern shaping. When $\upsilon = 1$ it reproduces DS-like behaviour and at $\upsilon = -1$ it approximates MVDR characteristics. The ability to adjust beamformer shape through υ makes FBF particularly valuable in conditions requiring balance between noise suppression, resolution and robustness. This is an effective enhancement tool for post-processing algorithms, like DAMAS, where the tunability helps shape the beamformed input pixel maps in order to balance the input.

7

Deconvolution Approach for the Mapping of Acoustic Sources

This is the fourth and last technique chapter and contrary to previous chapters this technique is not a beamformer. This is a post-processing algorithm called A Deconvolution Approach for the Mapping of Acoustic Sources (DAMAS) which uses beamformed pixel maps as its input. The chapter follows the same structure as the previous technique chapters.

7.1 Technique Description

The post-processing algorithm DAMAS is the fourth implemented technique. It operates on beamformed grid-based pixel maps in this setup. The DAMAS algorithm is an iterative process that "cleans" these grid-based maps by compensating for the microphone array's spatial response and reducing smearing effects. The goal is to refine the pixel-maps and better visualize the true source distribution which is concealed by mentioned effects in the beamformed maps.

The DAMAS algorithm was proposed by the NASA scientists Brooks and Humpphreys, Jr [30]. The purpose of the algorithm was to increase interpretability of array measurements and increase precision in source localization and strength estimation tasks in aeroacoustic applications.

The primary reason for implementing DAMAS as the fourth technique in this master's thesis is its ability to improve spatial resolution in situations where the previously implemented beamformer techniques alone struggles to separate sources. It is a well-established and standardized method within the research community, a substantial amount of literature has explored its potential and numerous adaptations of the original algorithm has been produced.

7.2 Theoretical Background

The DAMAS algorithm is an iterative process that takes a grid-based beamformed map as its input, together with the steering vectors used by the specific beamformer. The algorithm aims to resolve the true source distribution of measured sound field which is always obscured by artefacts and smearing effects caused by the arrays response. These effects do not belong to the true source distribution but their effect is always present in beamformed maps.

The theoretical background presented here is primarily based on Wu and Chang's work *Application of the Deconvolution Deblurring to Near-field Beamforming for Acoustic Array* [29]. It is also constructed upon the original DAMAS proposal by Brooks and Humphreys, Jr. [30]. The theory is adapted to suit an indoor pass-by measurement setup, taking into account unique constraints and leveraging properties of the specific system and environment.

7.2.1 Spatial Grid Discretization

The grid-based input maps represents spatially discretized source area of interest, where the region is divided into positions and pixels, illustrated in figure 7.1 showing both 1D and 2D setups. The black rectangle visualizes the area of interest, the gray lines depict grid divisions and black dots indicate the exact position where the beamformer is applied. The map illustrates the resolution of the acoustic image, emphasizing the importance of acoustic pixel-resolution in relationship to the desired spatial resolution.

The 2D grid-map is straightforward to understand and interpret, its looking down from the roof of the room towards the floor and discretizing an area where a source could be placed. The 1D grid-map is more complex, since our setup employs ULAs the DAMAS algorithm can have problems solving dimensions without microphone distribution. Therefore representing the data in 1D mode, or "slits" can be preferable. The approach ensures that the information presented to the user is is relevant and does not obfuscate interpretation with data that is practically non-accessible.

To better understand what these 1D slits represents, consider the thought experiment, *What would one-dimensional eyes see?* Imagine your eyes located in the same horizontal plane as the ULAs. If your vision was reduced to one dimension, your eyes would probably only be able to resolve images along the same horizontal plane as they are located in. In this dimension-deficient situation the vertical direction would not be resolvable as you have no extra eyes there. This has the implication that the horizontal plane would be dividable into portions, while no vertical resolution is available. Thus, creating 1D slits, areas which are horizontally divided but vertically infinite.

In relationship to the 2D grid-map representation, the 1D slits could also be thought of as distance based slices of the 2D grid-map, looking at just one distance at a time.

Figure 7.1: Illustration of grid mapping in 1D and 2D.

7.2.2 Model Setup

The DAMAS algorithm can be represented as an inverse problem where the goal is to iteratively estimate the true source distribution based on a beamformed grid-based pixel map. The beamformed map inherently contains smearing and imperfections caused by the microphone arrays response which acts as a spatial filter applied to the true source distribution. The problem can be stated in matrix form as:

$$\boldsymbol{Y} = \boldsymbol{A}\boldsymbol{X} \tag{7.1}$$

Where $\boldsymbol{Y}$ is the beamformed pixel-maps, $\boldsymbol{X}$ is the true source distribution and $\boldsymbol{A}$ is the DAMAS matrix, which characterizes the arrays spatial response. The goal is to solve for $\boldsymbol{X}$ using an iterative inverse estimation method. $\boldsymbol{Y}$ is known since it is the beamformed pixel-maps and $\boldsymbol{A}$ has to be calculated in order to estimate the arrays spatial filtering effect.

7.2.3 DAMAS matrix

The DAMAS matrix $\boldsymbol{A}$ aims to estimates the microphone arrays spatial response. If $\boldsymbol{Y}$ is the recorded microphone signal and $\boldsymbol{X}$ is the true source distribution, then $\boldsymbol{A}$ contains the array response which is convolved with the true source distribution to produce $\boldsymbol{Y}$. The matrix can be expressed as:

$$\boldsymbol{A} = \frac{|\boldsymbol{W}^H \boldsymbol{E}|^{\circ 2}}{M^2} \tag{7.2}$$

Where $\boldsymbol{W}$ is the weights related to a specific beamformer technique and it has the shape microphones x angles. The $\boldsymbol{E}$ matrix contains steering vectors for that specific beamformer in the same shape, microphones x angles. The normalization factor M is total number of microphones in the array. The $^{\circ 2}$ denotes elementwise power of two. Taking the Hermitian of $\boldsymbol{W}$ ensures that the DAMAS matrix A has the of angles x angles which corresponds too the total number of pixels at one certain distance.

For the $\boldsymbol{A}$ matrix to be suitable for the iterative solver (presented in the next subsection 7.2.4) it needs to be normalized. Due to the formulation of the matrix, all elements are

positive and the diagonal values contains the largest value present in the matrix. This allows for simple diagonal maximum value normalization. The matrix is strictly not diagonally dominant, since it requires the diagonal element of a certain row to be larger than the sum off the other elements in that row. It still exhibits diagonally dominant elements, in the sense that diagonal elements are the largest values present in the overall matrix.

7.2.4 Iterative Solution

The $\boldsymbol{A}$ is often ill-conditioned and thus non-invertible, this poses a problem as the DAMAS model is set as an inverse problem, as seen in see equation 7.1. The solution for this is to use an iterative solver method for linear equation systems called Gauss-Seidel method (GSM) which solves the problem by estimating $\boldsymbol{X}$ and at the same time avoiding performing the actual inversion of $\boldsymbol{A}$.

The DAMAS implementation has one crucial deviation compared to standard GSM, it forces a positivity constraint per iteration, if the result from a specific iteration is negative, than the algorithms omits this result and sets it as 0. The DAMAS positivity constrained GSM solver can be stated as:

$$x_n^i = \max\left(0, y_n - \left(\sum_{n'=1}^{n-1} a_{nn'} x_{n'}^i + \sum_{n'=n+1}^{N} a_{nn'} x_{n'}^{i-1}\right)\right) \tag{7.3}$$

This equation is stated for a single iteration i at pixel level n for all N pixels. This means that x_n, and $x_{n'}$ are scalars, referring to specific elements of the true source distribution matrix $\boldsymbol{X}$. Similarly for y_n which is elements of beamformed map matrix $\boldsymbol{Y}$ and $a_{nn'}$ are elements from the DAMAS matrix A. At the boundary cases $n = 1$ and $n = N$ the first or second summation terms becomes 0 due to indexing. During the first iteration $x_{n'}^{i-1}$ is initialized as the corresponding pixel values of $\boldsymbol{Y}$.

One DAMAS iteration update all pixels sequentially in order $n = 1, 2, ..., N$, followed by a reverse pass in order $n = N, N-1, ..., 1$. This is a complete DAMAS iteration where the algorithm goes forward and backwards through every pixel N.

So what the algorithm essentially does when it iterates over a specific pixel n is that it tries to estimate how much energy in pixel y_n can be attributed to surrounding pixels. Through the summation terms it estimates how much energy is contributed from previously calculated pixels in the same iteration $\sum_{n'=1}^{n-1} a_{nn'} x_{n'}^i$, and from remaining pixels, but from the previous iteration $\sum_{n'=n+1}^{N} a_{nn'} x_{n'}^{i-1}$. The contributions are subtracted from y_n and resulting x_n should contain less smearing due to array response than previous iterations.

7.3 Implementation

The implementation of DAMAS is different from the other beamforming techniques, As this is a purely post-processing algorithm. It uses beamformed pixel maps as input. The beamformer pixel maps coordinate systems is created by the `gridBuilder` function, described in

subsection 3.2.1. These coordinate systems are passed into the beamformers for evaluation.

Figure 7.2 shows a visualized flowchart of the implemented DAMAS algorithm. The **Input** step takes a previously described pixel map, expressed in a per frequency narrowband mode, in a 3D format with distances x angles x frequencies where distances and angles makes up the pixels, while the individual slices represents one frequency each. The corresponding steering vectors is inputted in a 4D format, microphones x angles x distances x frequencies. The last inputs needed is total number of iterations scalar and a percentage tolerance value which states that if no pixels has changed more than the percentage value between two iterations, the map is seemed stable and the algorithm breaks.

The **DAMAS Matrix** is calculated as described in equation 7.2. This equation uses weights $\boldsymbol{W}$ as an input. This DAMAS implementation applies the FBF as its input beamformer, it has shown superior performance and reliability compared to other beamformers. The applied $\boldsymbol{W}$ here is the steering vectors of the FBF, $\boldsymbol{W} = E^H$ as Wu and Chang proposes [29].

The larger dashed box contains the DAMAS adapted GSM solver. The **Iteration** step just denotes the start of an iteration and how the data moves throughout this module. The smaller dashed box represents everything that is part of a DAMAS iteration. Within this DAMAS iteration box, **Forward Pixel Order** highlights that the pixel map data is iterated in the direction of its original input sorting order. The pixel map is then put into the **Solver** where DAMAS calculates the actual GSM algorithm with positivity constraint, as presented in equation 7.3 where it estimates the true source distribution map $\boldsymbol{X}$. The next step **Backward Pixel Order** iterates over all pixels again but this time in reverse compared to original input pixel order. This forward/backward iteration can sometimes reduce reoccurring smearing patterns in $\boldsymbol{X}$.

At this stage one DAMAS iteration is performed and the algorithm checks if **Max iterations Reached?** is fulfilled based on both the maximum number of iteration input but also the percentage based tolerance value. The true source distribution map X is calculated per frequency and in the **Output** step the individual frequency maps are merged creating both a broadband and a narrowband per frequency output.

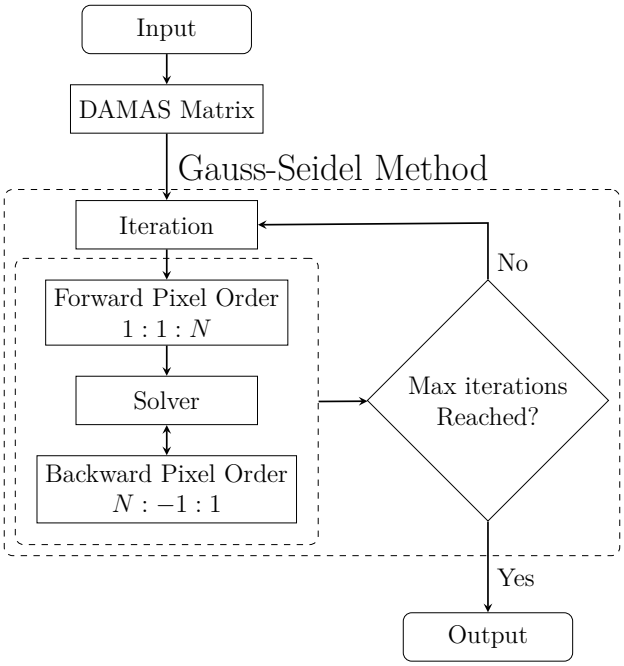


Figure 7.2: DAMAS function flowchart.

7.3.1 Output Pixel Maps

In the DAMAS theoretical background section, 1D and 2D output was described. Where it was stated that DAMAS has a problem estimating the true source distribution map $\boldsymbol{X}$ in dimensions where the ULAs has no microphone distribution. A solution for this was proposed where the data was presented in a 1D slits representation. This approach has the advantage of not presenting misleading data due to the algorithms problems of solving the additional dimension. Simplifying interpretation for users who are not aware of this deficiency. This section aims to present both a 2D grid-based map and 1D slits of the same simulation to visualize and explain the differences.

In figure 7.3 below, a grid and slit setup is illustrated. The left side plot shows the room setup with source locations and beamformer pixel locations in the semi-anechoic chambers Cartesian coordinate system. The central plots shows the FBF beamformer output with order factor $\upsilon = 2$. The right hand plots is the DAMAS output maps. In this setup the normalized dB tolerance is set to -32 dB from 0 peak dB, this could be reduced in order to achieve better source localization for the FBF beamformer in itself, but for the purpose of this demonstration and highlighting the arrays spatial response, it is set like this.

What these plots highlights is DAMAS problems solving the true source distribution in the y-direction. The sources are smeared and its representation are similar to lines, even though the real sources are monopoles. With the knowledge of this y-direction smearing it is still possible to use 2D-grid representation and based on visual density and intuition approximately locate sources in 2D. There are also situations where a 2D representation performs well but it is mainly due to the quality of the beamformed pixel maps input and not DAMAS solving-ability. The slit setup offers a simpler analysis tool which is easier to interpret and doesn't

require the same amount of knowledge from the user.

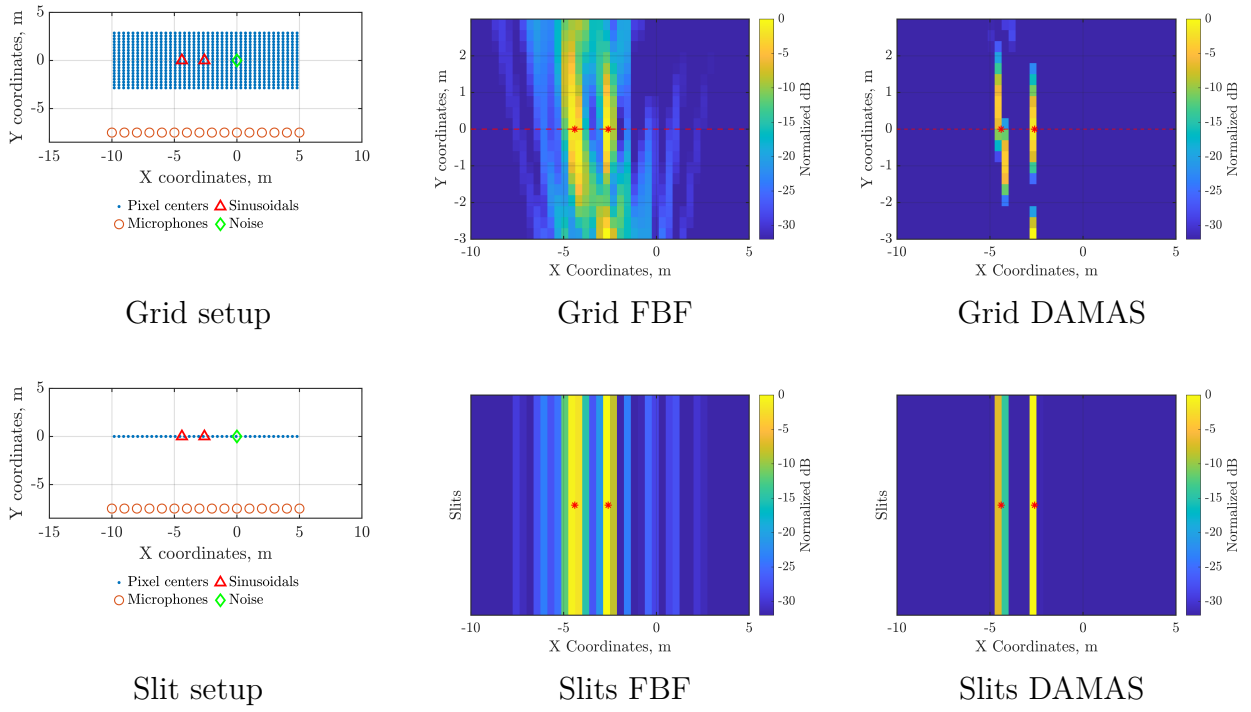


Figure 7.3: Grid and slits representation of FBF and DAMAS. Two sources at $f = 200,\ 300$ Hz left to right and noise source at $P = 0.1$ W. Red starts denotes True source position, red dashed line in grid plots shows at which distance the slit is placed.

7.4 Validation Metrics

This subsection presents validation metrics for the DAMAS post-processing algorithm. Specifically for DAMAS these metrics are custom designed. Since it is not a beamforming technique, its performance cannot be evaluated using the same set of validation metrics. DAMAS is also dependent on beamformed pixel maps as input, so its behaviour cannot be evaluated independently and its result is always impacted by beamformer performance.

The original reasoning for using DAMAS as a technique was to examine if it could improve spatial resolution and clean SRPs which contains high amount of smearing and artefacts. Therefore, the validation metric estimation will focus specifically on spatial resolution. The previously estimated validation metrics which are excluded here are inherently connected to beamformer techniques, and while DAMAS is dependent on beamformers for input, its performance is never better then the quality of the input map it receives, therefore it is important to ensure that the beamformed pixels maps are as high quality as possible.

7.4.1 Spatial Resolution

The spatial resolution estimation for DAMAS is based on the spatial resolution results from the previous beamformer techniques, see table 4.1. It is difficult for beamformers to separate

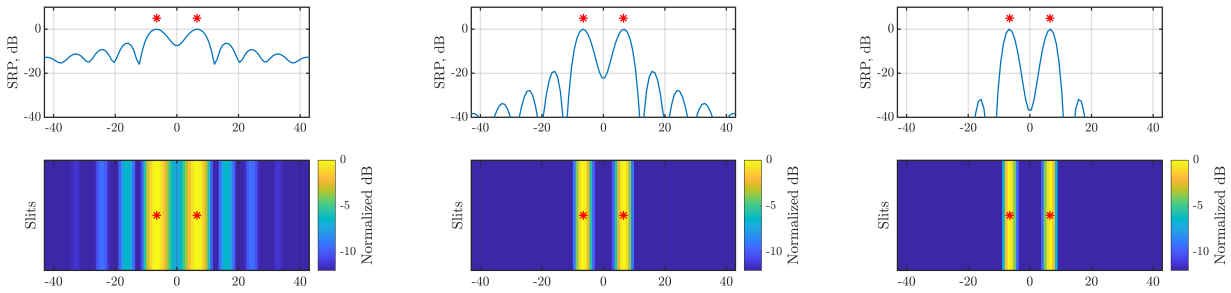
sources that are less then one wavelength λ apart and the ridge line between crosses and checkmarks is consistent with the individual frequencies λ. This analysis of spatial resolution using DAMAS will focus on λ and sub-λ distances between two sources at the same frequency, using the same spatial resolution setup as before.

The FBF will be used to produce the input pixel maps since the order factor υ offers valuable control of the lobe width and side-lobe suppression adjustment capabilities. Utilising FBF with positive υ has had the consistently best and most reliable performance with DAMAS.

The simulation environment first iteration is 2 sources placed at the physical distance matching their λ, if the DAMAS output map is deemed to be sufficiently cleaned, source-source distance is decreased in 10 cm increments. The DAMAS algorithm is iterative in itself and for these calculations a maximum limit has been posed at 100 iterations. The second demand states that if no pixels of the map has changed more than 1 ‰ compared to the previous iteration, the map is deemed stabilized and breaks. Generally, few trials in this setup see much improvement after 100 iterations, this is also impacted by grid sizes. In these simulations slits mode is used with 100 slits, each slit has an approximate width of 15 cm. In each simulation run the goal is that the map should have stabilized itself during the 100 iterations, if not υ of the FBF is adjusted. A simulation run for a specific frequency always start with high υ value, which is consistent with a narrower SRP, allowing for more precise source localization. If DAMAS fails to clean a map, lower υ value with wider mainlobe is tested, these maps are generally easier to clean but does not produce the same level of precision.

Interpretations of these maps has been hard to automatize, it is often a question of assessment if sources are sufficiently solved and sufficiently correctly placed. This has thus been done through a labour-intensive manual process where 1/3-octave band center frequencies has been tested. The figures has been carefully curated to highlight certain interesting behaviour of the DAMAS algorithm and its interplay with FBF. There is an additional table at the end which presented all tested 1/3-octave band center frequencies which is an example of what improvement DAMAS brings.

The figure 7.4 shows the FBFs array pattern and its slits representation. This is aimed to visualize the more standard SRP presentation type besides its slit representation. The source frequency is at 200 Hz, source-source distance is 1 λ and their true position is highlighted by red asterisks. υ is successively increased from plot **(a)** to **(c)**. This shows the effect of υ on the SRP. This is a very useful effect in combination with DAMAS as it gives an extra tuneable control of the input, being able to both narrow and widening beamformed pixel map input SRP.



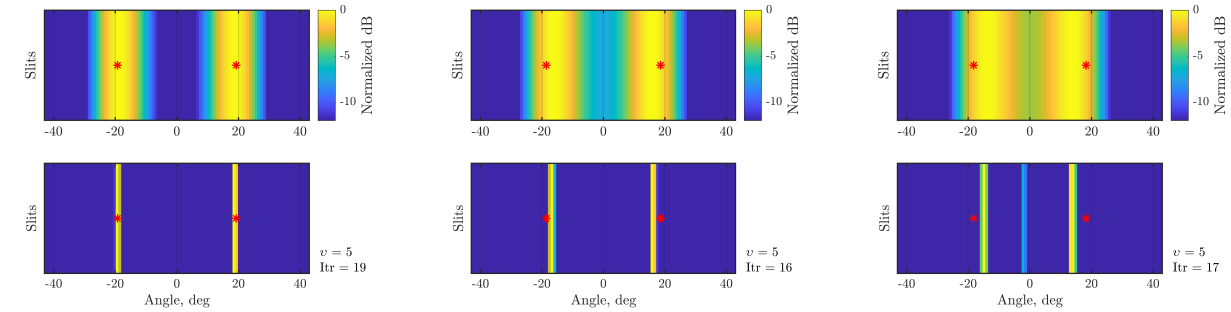
(a) $\upsilon = 1$, DS-CSM behaviour.

(b) $\upsilon = 3$.

(c) $\upsilon = 5$.

Figure 7.4: FBF SRP and slit mode SRP. Frequnecy 200 Hz at distance λ, different υ's. Red asterisks represent true source locations.

The figure 7.5 presents 63 Hz at 3 different distances relative to its λ, top figure is FBF SRP in slits representation and bottom plot is DAMAS output cleaned beamformed map. The plots highlight how DAMAS solvability decreases as source-source distances decreases while υ is kept constant. Both **(a)** and **(b)** are both classified as resolved and while **(c)** resolves the sources, the offsets versus the true source location is deemed to be to large.



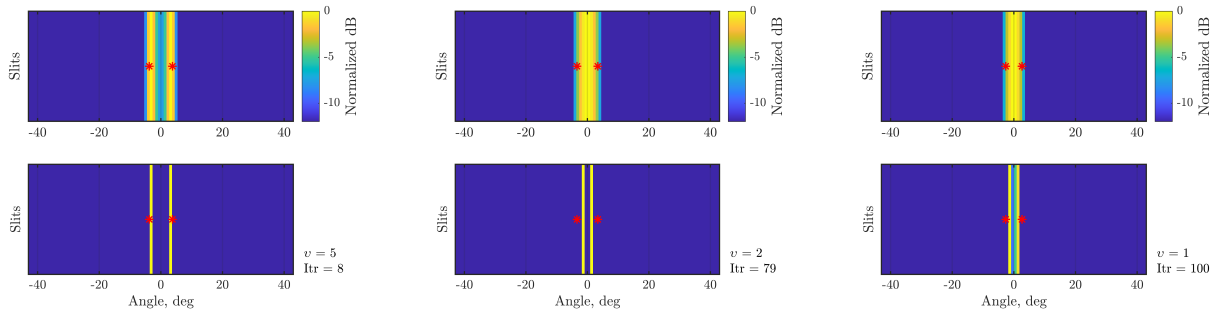
(a) $\lambda - 20$ cm.

(b) $\lambda - 40$ cm.

(c) $\lambda - 50$ cm.

Figure 7.5: FBF SRP and DAMAS output map. Frequency 63 Hz at $\upsilon = 5$, varying source-source distances.

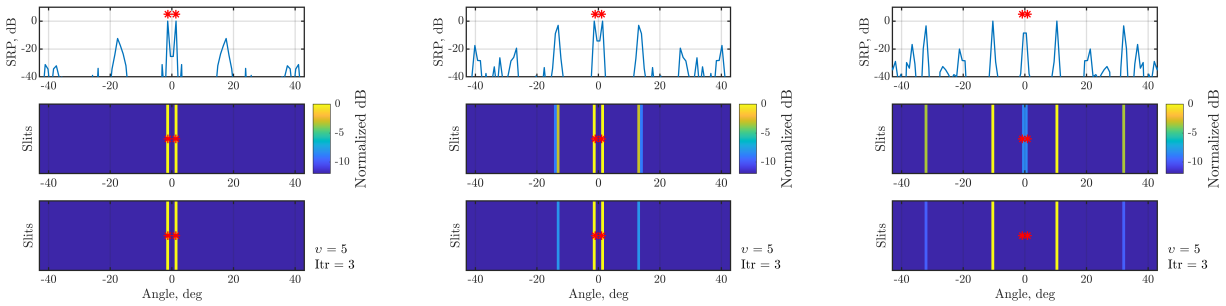
Next figure 7.6 shows two resolved DAMAS maps, **(a)** and **(b)** while **(c)** is on the brink of being resolved but falls into the category of unresolved, the sources are deemed to be not sufficiently spaced. These plots has falling υ values and depicts a pattern where smaller source-source distances often has to have lower υ to be resolved. Subfigure **(c)** is the first plot which iterates up until maximum 100 iterations without stabilisation. Tests where made relaxing maximum iteration demand, just keeping the 1 ‰ demand, the total number of iterations ended close to 200 iterations but no visible changes could be seen in the result. Further changing maximum iterations to 500 without the 1 ‰ demand did not improve the result either.



(a) $\lambda - 10$ cm. **(b)** $\lambda - 20$ cm. **(c)** $\lambda - 40$ cm.

Figure 7.6: FBF SRP and DAMAS output map. Frequency 315 Hz at varying v and source-source distances.

Comparing all the previous figures in this chapter, it is possible to draw the conclusion that DAMAS needs some spacing in the SRP in order to successfully solve a source distribution. At higher frequencies, DAMAS solvability at sub-λ distances diminishes at higher frequencies but it can successfully solve two λ spaced sources at 1250 Hz, see figure 7.7. At 1600 Hz grating lobes has overtaken the main peaks as seen in the SRP and thus DAMAS cannot solve for the correct source distribution.



(a) $f = 1000$ Hz. **(b)** $f = 1250$ Hz. **(c)** $f = 1600$ Hz.

Figure 7.7: FBF SRP and DAMAS output map. Varying frequencies at same distance λ and $v = 5$.

The table below shows 1/3-octave band center frequencies for all valid frequency bands and the result of the DAMAS output maps. Higher frequencies not part of this table but still within frequency range of interest are not plotted. From 400 Hz to 1250 Hz DAMAS could only solve source-source distance λ, higher frequencies than 1250 Hz was non-solvable. Green boxes means that DAMAS successfully separated sources with sufficient spatial accuracy. Each box contains values for v/iterations. Center frequencies above 500 Hz are not showed. The algorithm is most effective at lower frequencies, achieving a sub-λ source separation at -40 cm for wavelengths larger than 80 Hz.

Hz \ λ−cm	$\lambda - 40$	$\lambda - 30$	$\lambda - 20$	$\lambda - 10$	λ
40	5/12	4/100	5/15	5/18	5/14
50	5/14	5/15	5/16	5/16	5/19
63	5/16	5/17	5/19	5/19	5/21
80	5/15	4/100	5/20	5/23	5/21
100	✗	5/23	5/15	5/22	5/27
125	✗	5/23	5/14	5/19	5/19
160	✗	✗	5/19	5/10	5/28
200	✗	✗	2/78	5/8	5/8
250	✗	✗	4/100	5/8	5/14
315	✗	1/100	2/79	5/8	5/4
400	✗	✗	✗	5/7	5/8

Table 7.1: Spatial resolution, each box contains values υ/iteration.

7.5 Analysis

The figures 7.5, 7.6 and 7.7 all showed a pattern where approximately a 6 dB dip from maximum is needed between two sources for DAMAS to reliably resolve two sources. DAMAS is generally able to solve lower dB-differences also but at the cost of reduced source localization precision, which is visible in figure 7.5 where at distance $\lambda - 20$ cm, sources are spot on true source position. As distances decreases DAMAS source localization migrates towards the center compared to true source potion, overshooting the true position.

With DAMAS, sub-λ distance precision can be achieved at lower frequencies but the improvement generally diminished as frequency increases, see table 7.1. The inter-microphone distance between these ULAs are 1 m and DAMAS results are reduced drastically for $\lambda < 1$ m if this is a coincidence or actually connected to the DAMAS algorithm remains unclear.

The DAMAS algorithm is very dependent on the quality of the beamformed map inputs and its the single most important factor for achieving reliable, high precision results. This is also why its potential is realized in combination with FBF, since this beamformer offers array pattern tunability through its order factor υ. If high precision source localization is the goal, than a higher υ with narrower mainlobe is generally the preferable input. If robustness is prioritized, then lower υ with wider mainlobe and lower side-lobe suppression is better.

This is the last implemented technique so there is no need for a justification or recommendation for a next technique to implement. DAMAS achieved better spatial resolution than what previously methods has been able to provide, but the difference of it is not very large. Comparing the results from table 7.1 to table 4.1 wouldn't tip any 1/3-octave bands into green checkmark territory but this is of course due to the coarseness of the range resolution of the second table. Generally the basic form of the FBF beamformer, DS-CSM coupled with DAMAS is able to source separate frequencies at 1 λ distance from 50 Hz to 1600 Hz. Grating lobes appearing at 1600 Hz is a deficiency of the FBF beamformer and the ULA

configuration and not an effect of DAMAS, but since DAMAS is dependent on beamformer input, its ability is not possbile to estimate by itself.

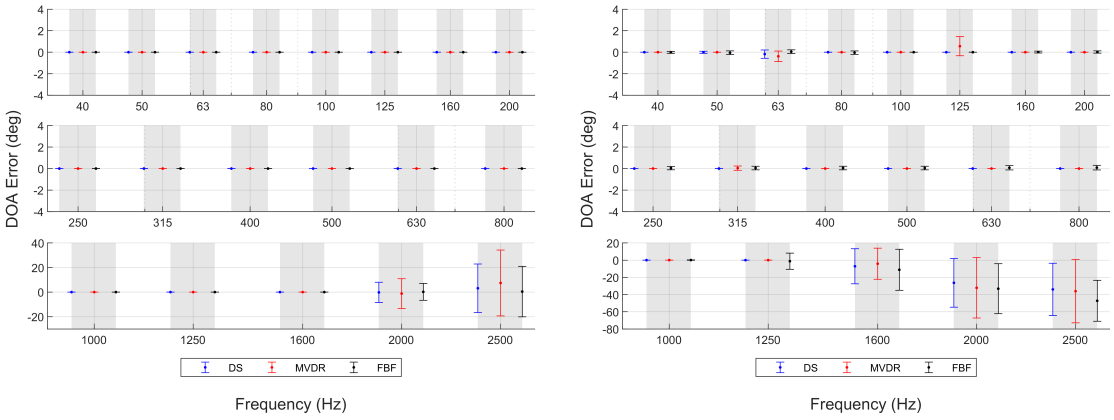
8

Comparative Results

This chapter presents a comparative analysis of the individual techniques, evaluating their performance and limitations in scenarios relevant to the objectives and aims of this master's thesis. The analysis is based on measured pulse train and noise data recorded in the semi-anechoic chamber, together with the technique-specific validation metrics presented in the individual techniques chapters. Finally, all techniques is tested on real indoor pass-by measurement data provided by Volvo. The result should highlight the suitability of each technique in different situations and may also serve as a practical guide for the user of the acoustic camera software.

8.1 Validation Metrics Comparison

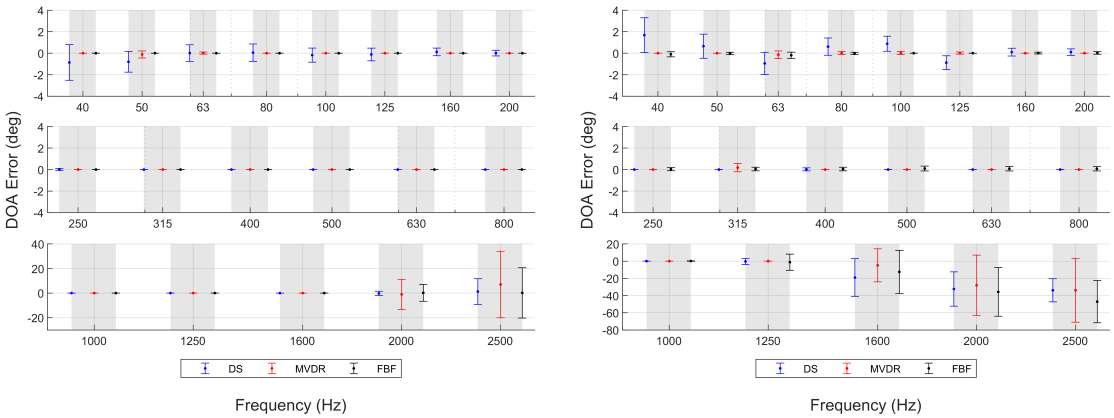
Figure 8.1 presents a comparison of DOA estimation errors for the DS, MVDR, and FBF beamformers under ideal and noise-free conditions. Across both on-axis and off-axis configurations, all three techniques demonstrate consistent performance with relatively low estimation errors up to 1600 Hz. The DS beamformer shows slightly better accuracy over whole frequency band compared to the other techniques. The similarity in performance across techniques under ideal conditions provides a useful reference point for later evaluating robustness under more challenging scenarios.



(a) On-axis DOA estimation error. **(b)** Off-axis DOA estimation error.

Figure 8.1: Comparison of DOA estimation error for DS, MVDR and FBF under ideal conditions. **(a)** shows the results for an off-axis source position, **(b)** shows the results for an on-axis configuration.

Figure 8.2 compare DOA estimation errors across DS, MVDR and FBF under noisy conditions and comparing both on-axis and off-axis source positions. In the low frequency range (40 Hz - 200 Hz) the DS beamformer exhibits a notably larger spread in estimation error. Indicating high sensitivity to noise. In contrast MVDR and FBF maintain significantly more stable performance in this region. In the mid and high frequency region up to 1600 Hz for on-axis sources. In the off-axis configuration estimation errors begin to spread more around 1250 Hz, reflecting the decreasing spatial resolution and growing mismatch between actual and assumed wavefront geometries.



(a) On-axis DOA estimation error. **(b)** Off-axis DOA estimation error.

Figure 8.2: Comparison of DOA estimation error for DS, MVDR and FBF under noisy conditions. **(a)** shows the results for an off-axis source position, **(b)** shows the results for an on-axis configuration.

To asses technique robustness to microphone sensitivity mismatch, a 20% random gain offset was introduced to each channel. This level of offset is considered high, even a 5% deviation would result in re-calibration. Despite this the effect on DOA estimation accuracy was negligible across all beamformers. Although such mismatches are expected to influence the structure of the CSM, its impact appears minimal in the evaluated conditions.

The valid frequency ranges for each method is summarized in Table 8.1. This shows that all beamformers maintan accurate DOA estimation up to 1600 Hz on-axis while performance for the off-axis case degrades slightly around 1250 Hz.

Technique		Noiseless, Hz	Noise, Hz
DS	On-Axis	40 - 1600	40 - 1600
	Off-Axis	40 - 1250	40 - 1250
MVDR	On-Axis	40 - 1600	40 - 1600
	Off-Axis	40 - 1250	40 - 1250
FBF	On-Axis	40 - 1600	40 - 1600
	Off-Axis	40 - 1250	40 - 1250
ALL	On-Axis	40 - 1600	40 - 1600
	Off-Axis	40 - 1250	40 - 1250

Table 8.1: Valid beamformer frequency ranges based on validation metrics.

In summary, the comparative evaluation shows that MVDR and FBF offer improved robustness and consistency under noisy conditions, particularly at low frequencies. However, they are more sensitive to microphone placement offsets compared to DS, which remains more tolerant to geometric deviations despite its higher noise sensitivity. All methods operate reliably up to 1600 Hz on-axis and slightly lower off-axis with moderate sensitivity mismatch. These results support the use of MVDR and FBF in applications where array imperfections and environmental noise are expected.

8.2 Pulse Train Measurements

The pulse train measurements are used to assess the beamformers valid frequency range in a real-world scenario. The pulse train consists of sinusoidal tones at 1/3-octave band center frequencies in the frequency range 50 Hz to 2 kHz with some silent spacing between each tone, subsection 3.3.1 thoroughly explain the signal types and the their purpose in this setting.

Reiterating some relevant parts from the aims and purpose sections of this master's thesis. The main focus is source localization of distinct tonal or similar sources within a possible source area in the frequency range 50 Hz to 2 kHz. The pulse train signal contains one extra 1/3-octave band center frequency below and above defined frequency range.

The presented pulse train measurements originates from two different measurement sessions, both containing two source locations each. One position centralized in front of the array, the second location at the furthest edge of the possible source area. Measurement sessions were conducted at different distances from the microphone array. See figure 8.3 which presents the setup. Position 1 and 2 are from the same measurement session. Position 3 and 4 are from the second session. The noise source was used in conjunction with position 3 and 4. The blue dots represents beamforming pixels for each source distance.

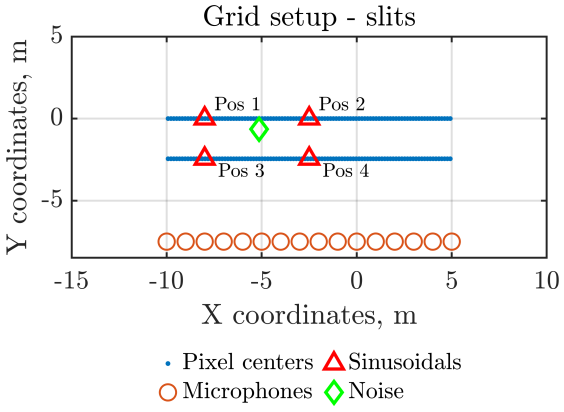


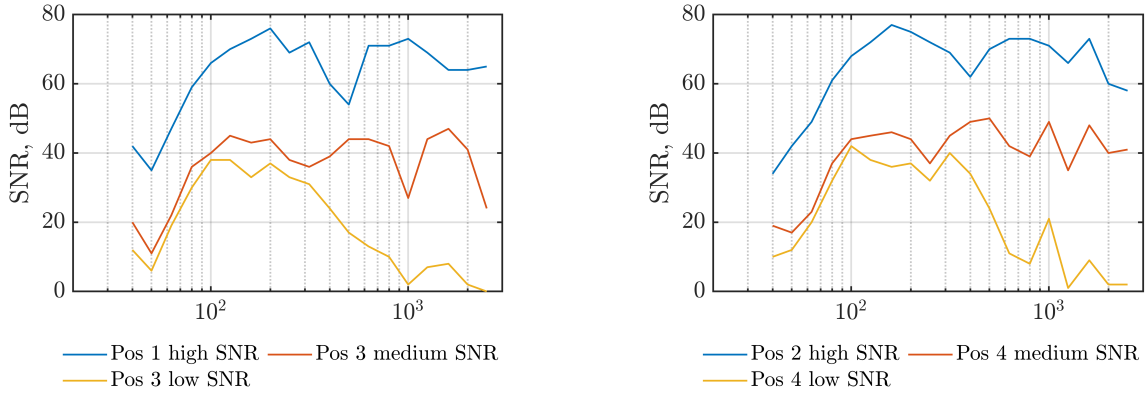
Figure 8.3: Location and grid setup for measurement sessions 1 and 2.

Pos	SNR		
	Low	Medium	High
1			✔
2			✔
3	✔	✔	
4	✔	✔	

Table 8.2: SNR case for each source position.

Three different SNR cases is used in these measurements, not all positions present the same SNR cases, table 8.2 contains a summary of SNR cases presented per measurement position.

SNR is depicted for position 1 and 3, position 2 and 4 together in figure 8.4. The SNR is presented per 1/3-octave band where pulse and background signals has been filtered accordingly. This makes the SNR levels seem overall high but it is due to the limited frequency range being compared, if the signals were truly broadband, SNR would be lower due to additional noise in the spectral content which would not be a faithful comparison, this is explained more thoroughly in subsection 3.3.1.



(a) Position 1 and 3 **(b)** Position 3 and 4

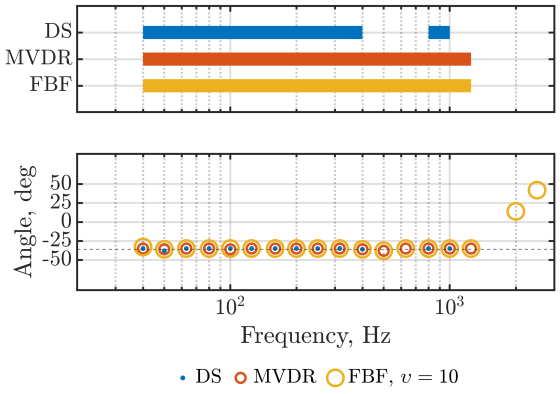
Figure 8.4: SNR per position and case.

At lower frequencies the SNR is generally worse, independent of SNR case, and this is likely due to a combination of frequency response for both the loudspeaker and the amplifier, and the background noise levels present in the semi-anechoic chamber. The degradation of SNR, especially at higher frequencies, in the worst SNR case is due to the additional noise source. Comparing SNR cases medium and low for position 3 and 4, the difference between them is the introduced noise source, which highlights that it does a better job masking the higher frequency range.

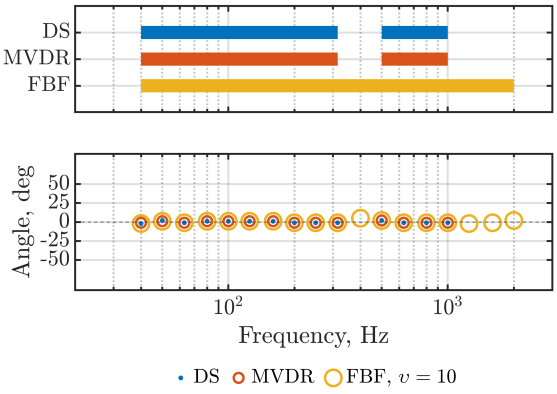
Upcoming figure 8.5 depict the individual beamformers ability to localize sources based on the pulse train measurements, positions and SNR cases presented here. Each position consists of two plots, one upper, which visualizes each beamformers valid performance range, the lower plot shows angle versus frequency and highlight the beamformers source localization ability.

In order to interpret these figures correctly, it is important to have certain prerequisites. The lower figure visualizes each beamformers ability to localize a source angle compared to the true source angle. If a certain beamformers circle is missing at a specific 1/3-octave frequency band it means that the SRP had more than one peak above the tolerance value -3 dB on a normalized dB scale and its SRP was deemed not suitable for source localization purposes. The upper performance plots depicts the valid frequency ranges of the beamformers, they use an additional angle tolerance value, where a located source should be within 10° distance from the true source location in order to be perceived as a correct estimation. If an estimation is within both permitted tolerance values for peaks and dB-level, and there are at least two correctly source localized frequencies in a row, it is deemed to be in the valid frequency source

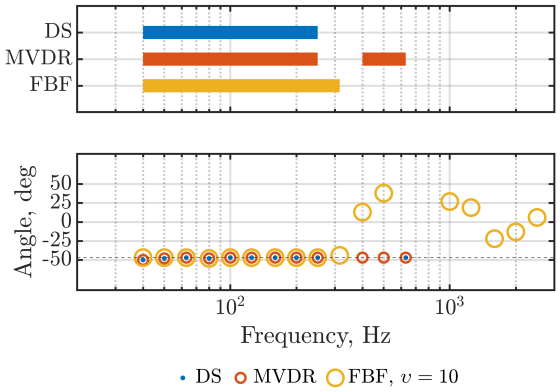
estimation range of a specific beamformer.



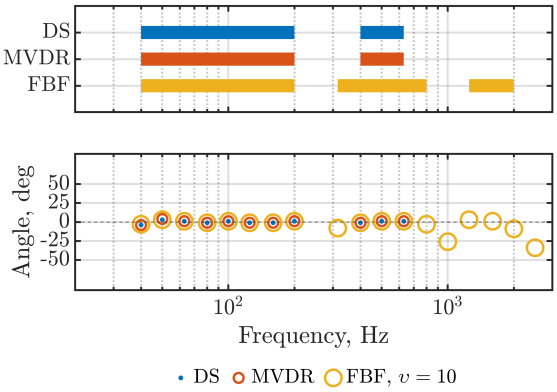
(a) Position 1, high SNR.



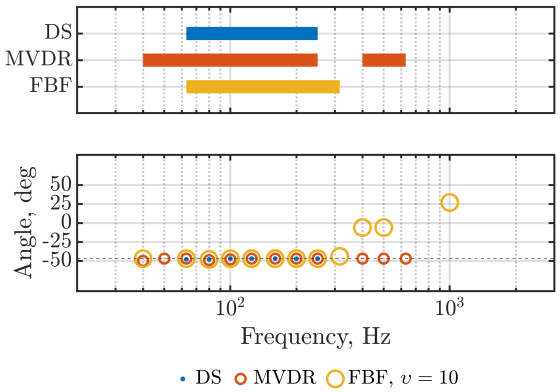
(b) Position 2, high SNR.



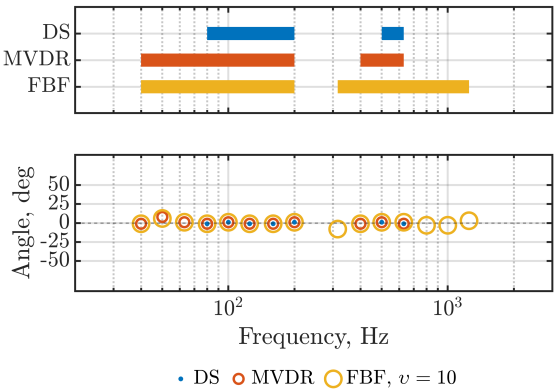
(c) Position 3, medium SNR.



(d) Position 4, medium SNR.



(e) Position 3, low SNR.



(f) Position 4, low SNR.

Figure 8.5: Beamformers pulse train measurements source localization performance.

There are some obvious conclusion which can be drawn from the plots. The beamformers performs well in low frequency range in a single source situation. At low frequencies the SRP mainlobe will be much wider and is not able to separate sources, see table 4.1. But since these measurements are single sources, the effect of wide mainlobe is not easily distinguishable. The SNR is very important for the beamformer source localization ability, which is expected since to low SNR impacts the signal part and masks it in noise, reducing it's importance as a distinct sound field component.

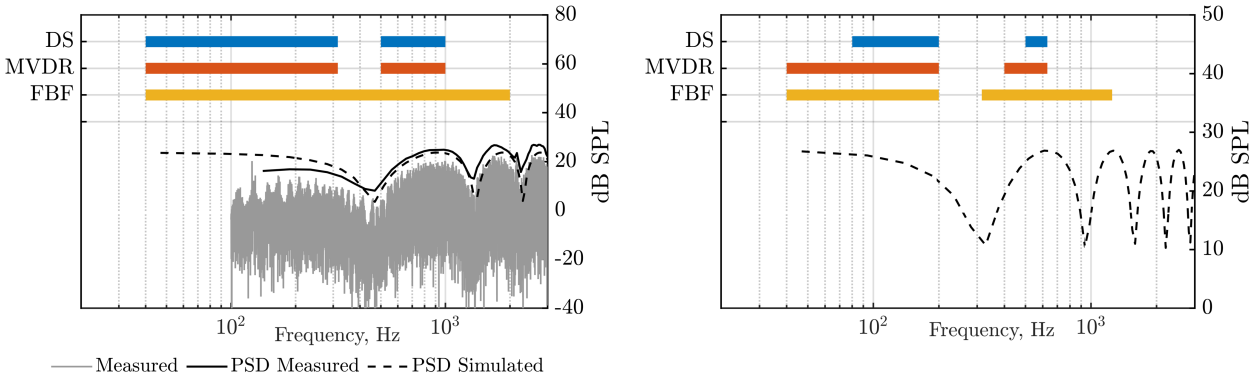
Analysing the valid frequency range from these pulse train measurements and comparing them to validation metrics valid frequency ranges, see table 8.1. The high SNR pulse train measurements achieves similar overall frequency range as the validation metrics. With the source placed on-axis high SNR case, see figure 8.5b, DS and MVDR performs 2 1/3-octave band center frequencies worse, achieving correct source localization at 1 kHz while FBF actually performs better than validation metric estimation reaching 2 kHz. Source placed off-Axis, figure 8.5a, MVDR and FBF reaches 1250 Hz which is the same as expected from validation metrics estimations.

As SNR decreases all beamformers source localization performance deteriorates as expected because the level differences between actual source and noise diminishes. Comparing low SNR case positions figures 8.5e and 8.5f with SNR plotted in figure 8.4, the low SNR case for position 3 and 4 shows deteriorating SNR from 200 Hz with occasional SNR increases at position 4. This temporary increases could explain why FBF manages to reach 1250 Hz at position 4 but just 300 Hz at position 3.

Analysing the valid frequency ranges for position 2 and 4, there is a reoccurring dip in the beamformers ability to source localize correctly. At position 2, which is at 7.5 m distance from the ULA, there is a dip visible in DS and MVDR at 400 Hz. At position 4 (5.05 m from the ULA) the frequency of the dip has decreased to 250 Hz, this also breaks the FBF. Comparing SNR case medium and low for position 4 seems to widen the dip when SNR decreases. The same dip is visible for position 1 and 3 but it is not as distinguished, at position 1 high SNR, only DS is affected and recovers at higher frequencies. At position 3 both medium and low SNR case only MVDR manages to recover.

The dip seems to shift frequency range by distance as there is no frequency location shift between position 4 medium and low SNR, just some widening which is expected because of worse SNR. But at position 2 high SNR, the dip is located at higher frequencies. Utilising white noise measurements recorded at position 2 and comparing this to the dip frequency location it is possible to draw the conclusion that the dip is due to destructive interference from the concrete floor reflection. Figure 8.6 plots valid beamformer frequency range for positions 2 high SNR, and position 4 low SNR (the same as in figures 8.5b and 8.5f) and compares the result to the measured noise signal, PSD of measured noise signal and PSD of simulated noise signal at the same location. The measured signals and the PSDs at position 2 in figure 8.6a is cut at 100 Hz due to uncertainties about the evenness of the noise source frequency response in combination with background noise levels. The simulated PSD signal frequency limit is set lower to highlight that the dip visible at 400 Hz is the first expected dip because of the destructive interference. At position 4, with source-ULA range 5.05 m,

figure 8.6b, no noise measurements were available and we relied on simulations in order to verify that the dip in valid beamformer frequency range is due to the floor reflection.



(a) Position 2, high SNR.

(b) Position 4, low SNR.

Figure 8.6: Beamformer valid frequency range versus measured and simulated floor reflection interference.

Visualized Source Localization

This subsection presents two different pulse train measurements at the same 1/3-octave band center frequency of 200 Hz. The purpose is to visualize each beamformers source localization ability using the pulse train measurement data. The axes in the figures are presented on a linear normalized scale rather than logarithmic, the lover values boost of the logarithmic scale is undesirable for source localization purposes.

Figure 8.7 presents two different measurement position, one per row. The first row is off-axis position 3, as presented in figure 8.3. The second row is an additional measurement position not previously presented. Each column of the figure represents one beamformer for DS, MVDR and FBF with $v = 10$. The data reflects SNR case low.

All positions related to the previously presented pulse train measurements was placed at approximately 1.4 m, as close as the microphone height at 1.2 m that the loudspeaker stand would permit. The additional measurement position (row 2) has source placed at 2.3 m height, considerably higher than the microphone plane.

These plots highlights each methods source localization ability when applied to the pulse train measurement data. Smearing effects are prominent in DS beamformer, but the MVDR and FBF results are similar to each other. MVDR had slightly better performance off-axis at position 1, and FBF performs slightly better at the additional position. All beamformers manages to localize the source correctly at both positions, close to them microphone plane and at an increased height.

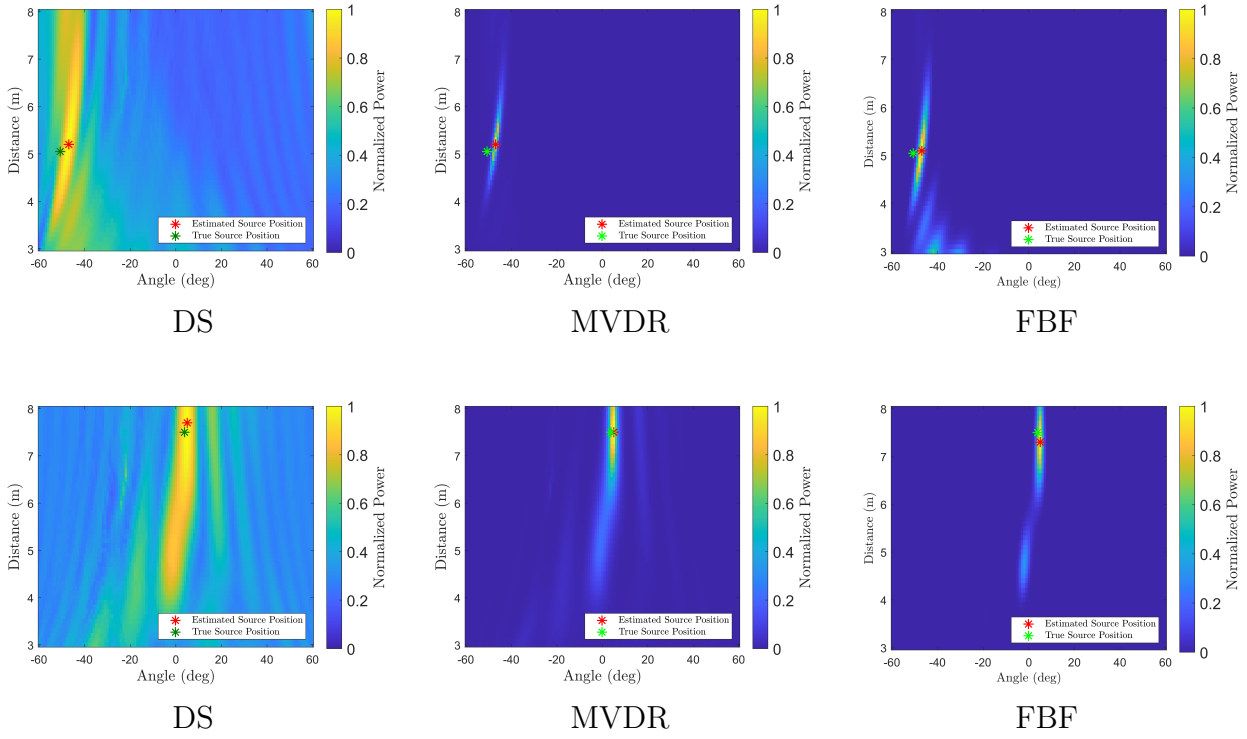


Figure 8.7: Pulse train measurement at frequency 200 Hz, source localized at two positions, position 3 and an additional not previously presented position. Per column each beamformer is presented, DS, MVDR, FBF with $\upsilon = 10$.

8.3 Vehicle Measurements

This section presents a practical test case of the acoustic camera software applied to vehicle measurement data. All 3 beamforming techniques and DAMAS is applied to the provided data with the goal of localizing a source of unknown origin.

The measurement data was provided by NVH engineers at Volvo and contains pass-by measurement recordings from within the semi-anechoic chamber. The measured vehicle is a truck and the timbre of it contains a tonal component and with the help of the acoustic camera software we will hopefully be able to identify the source location correctly. Since these are real vehicle measurements, the actual dB SPL values cannot be presented, they have been intentionally obscured or hidden for presentation purposes. The truck pictured in the upcoming plots is a generated image based on a Volvo truck, but it has nothing to do with the actual measured vehicle, it is used here as a substitute for visualization purposes. The side-view truck picture has been rendered using Volvo Truck's online Truck builder tool [31]. The from-above picture was generated using Volvo Truck builder images as prompt to OpenAIs GPT-5 [32].

The Volvo NVH engineers has on beforehand identified the noisy component and its location, however the information was not disclosed to us. At the start we had no more knowledge than what has been provided here, there is a tonal component in the measurements, and the

goal is to try to localize it. The following section reiterates the same process followed by us from receiving the measurement data, analysing using each techniques, the results and any decision taken during the analysing process.

Initial Analysis and Setup

The first step of the analysis was to investigate the data content, listening, trying to identify the tonal noise component and finding a time location where it is prominent. Figure 8.8 presents a magnitude spectrum of the individual microphone signals where the levels has been intentionally hidden from view. The visualized segment contains 1 s of data which was used as the input for all beamformers. At this stage the tonal component is visible in the range 900 Hz to 1 kHz. Previous use and analysis of the beamformers has revealed that they are all sensitive to low frequency noise and therefore a high-pass filter (HPF) with a cutoff frequency at 500 Hz has been applied, since no signal content below this frequency threshold will be off further use.

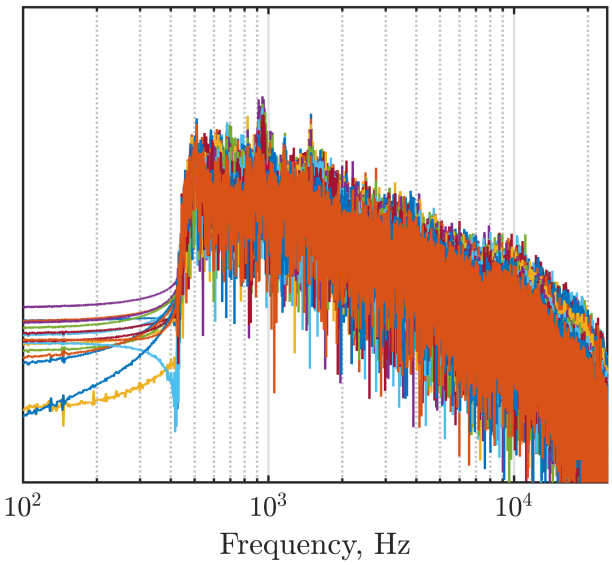
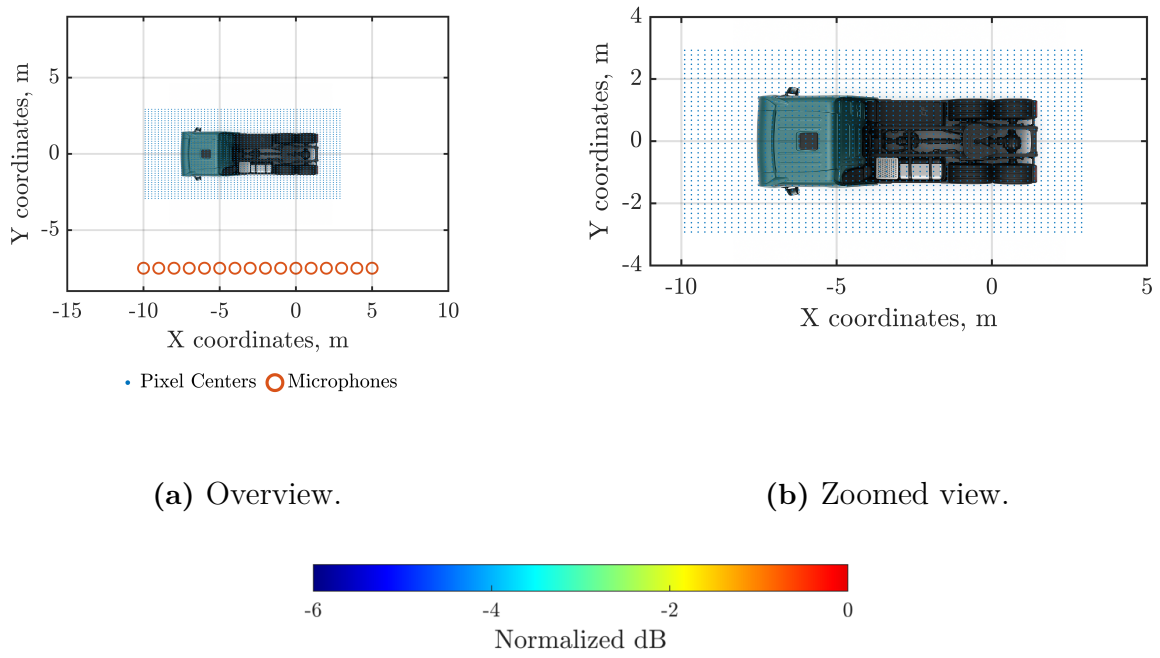


Figure 8.8: Location and grid setup for measurement sessions 1 and 2.

The analysis will primarily be performed in a 2D grid environment utilising the `gridBuilder` function except for DAMAS which will be examined in both 2D grid mode and 1D slits. Figure 8.9 depicts the room setup with microphones visible, right hand side plot is a zoomed in version which is the setup for further analysis. All blue dots represents a pixel center and one aiming location for the beamformers. It is a 60 x 40 pixels grid which produces an approximate pixel size of 330 cm^2 per pixel. X-axis resolution at 20 cm and y-axis at 15 cm per pixel. A normalized dB axis is displayed which is used for all upcoming beamformer plots. Each plot is normalized per the maximum occurring value in the pixel map and any values within 6 dB from this peak is visualized in the plot. All plots are presented in dB levels. As were mainly interested in source localization it could be beneficial to do presentation in linear scale which do not enhance lower value representation, but in this situation it was deemed

to hard to find a normalized scale which could handle all beamformers therefore a dB-scale representation was chosen anyway.



(a) Overview. **(b)** Zoomed view.

Figure 8.9: Room setup, zoomed view and normalized dB axis.

DS

The first technique to be tested was the initially implemented DS beamformer. Based on previous findings this technique was expected to perform the worst, but it finds a distinct target location placed centrally on the cabin. There is smearing pattern and array pattern artifacts in the -6 dB range but there is a distint peak closer to 0 dB which is undisputed.

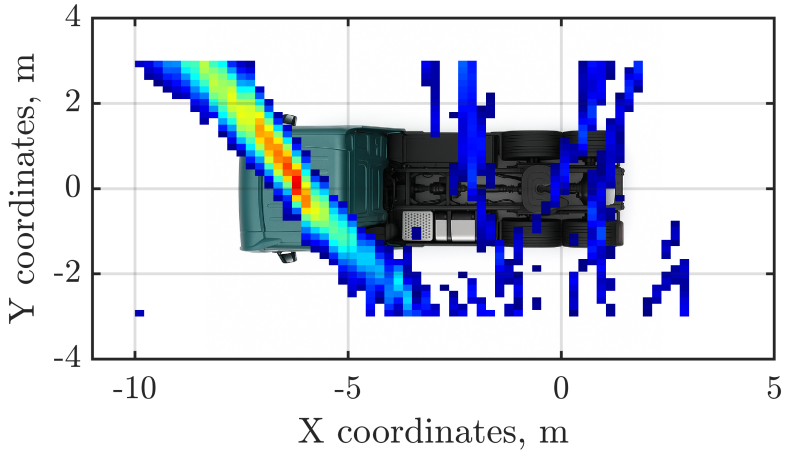


Figure 8.10: DS.

MVDR

The MVDR beamformer is expected to have a sharper peak than the DS beamformer and since the previous technique found the target, MVDR is expected to find a target at a similar location but with a narrower response due to the MVDR weighting. The MVDR is applied in the frequency range of 900 Hz to 1 kHz which means that it calculates its response for these specific frequencies and omits results from outside this range, essentially acting as a band-pass filter. The expected narrower array response is visible in the MVDR plot, where smearing pattern beyond the cabin has largely disappeared. There is some smearing still left which moves in a pattern toward the center location of the array and outwards, this is a typical smearing pattern for beamformers and is often distinguishable at varying levels in different results and situations.

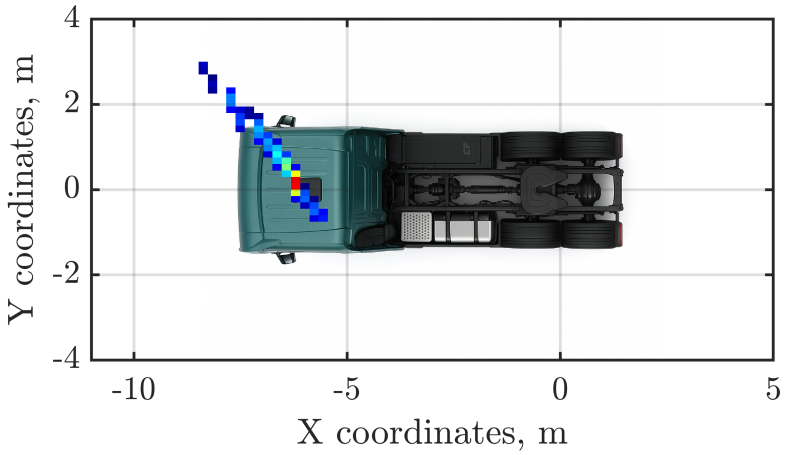


Figure 8.11: MVDR.

FBF

The FBF technique has an extra level of adaptability through the order factor v, which can be tweaked to achieve a satisfying result. Multiple v has been tested through trial and error to find optimum values. The FBF is band-limited to frequency range 900 Hz to 1000 Hz. Since both previously examined techniques found a target, FBF is expected to do so as well, with some tweaking of v the pattern is expected to contain less smearing than any of the previous techniques. 3 different values of v is tested and presented below, two positives and one negative.

The first positive values is $v = 5$, visible in the upper plot, which shows reduced smearing compared to previous techniques. Increased $v = 15$ (central plot) shows basically no smearing left at all. Due to this very exact ability to pinpoint source positions, there is some uncertainties of what level of accuracy can be expected in situations similar to this, when it is primarily the grid and pixel size which sets the resolution limitation. Experimental use of different grid sizes has shown that increasing number of pixels further than this doesn't necessarily enables increased source localization ability, just smears the same result over more pixels. There are 3 pixels with normalized dB values between 0 dB and -6 dB, this means that the source is localized to an area of approximate size 990 cm^2, or located in close vicinity.

The negative v is the first fluke of the vehicle measurements, it cannot localize the same source positions as the other beamformers, instead it places maximum just northwest of the cabin. Negative v has been shown to behave similarly to MVDR in terms of how its array pattern behaves, with the characteristic pointy peaks of MVDR, see figure 6.1 comparing the beamformers. The results of negative v FBF also tends to be similar to MVDR, but not in this case as MVDR manages to resolve a source location while negative v FBF does not.

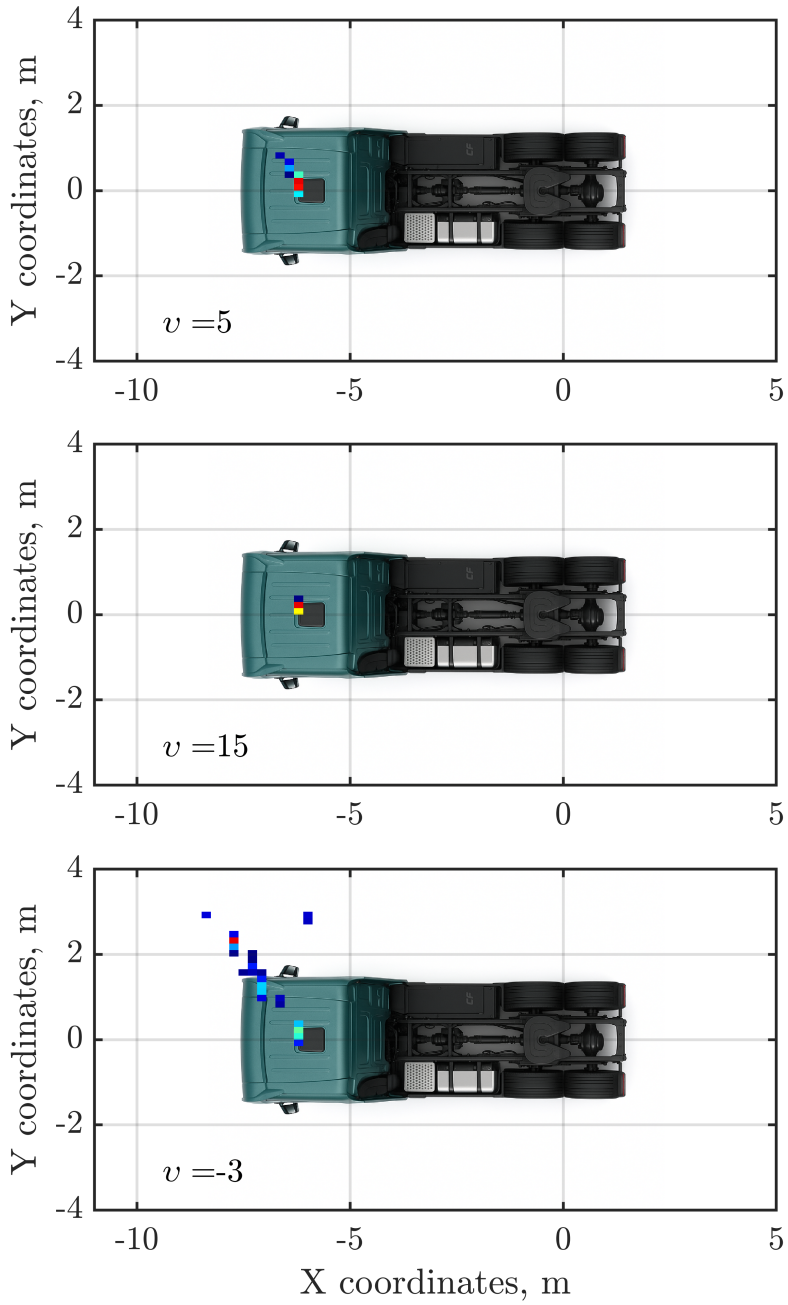


Figure 8.12: FBF with different v.

DAMAS

The DAMAS post-processing algorithm takes a grid-based beamformed input map and it aims to resolve the true source distribution. Because of this DAMAS is sensitive to the quality of the beamformed pixel map, if the map doesn't resolve any sources, DAMAS will not either be able to do so. However in the case where a map partially manages to resolve a source distribution, DAMAS can be applied in order achieve further separation.

The input to DAMAS is based on the FBF beamformer maps, but it is not using the best FBF $\upsilon = 15$, there is nothing left for DAMAS to resolve there, instead were using the $\upsilon = 5$ result. So the DAMAS figures should be compared with figure 8.12 upper plot. Generally, when υ is increased, the mainlobe narrows and side-lobe levels of the FBF are reduced but this does not always translates to stability for the DAMAS algorithm, instead it could be beneficial to use a slightly lower υ to achieve a more robust and reliable DAMAS result.

The DAMAS algorithm is an iterative process and takes a number of iterations as input, for analysis purposes it is beneficial to try both lower and higher number of iterations in order to understand the DAMAS behaviour when applied to this specific sound field. With the FBF $\upsilon = 5$ input the beamformed map is already considerably clean and just 5 iterations is sufficient to resolve the pattern enough.

Usually DAMAS only manages to solve a sound field in the same dimension as the microphones of the ULA is located within, this often leaves trace lines which obfuscates the result, see figure 7.3 which depicts how these lines can manifest. Here, the beamformed input pixel map is already solved to such an degree that DAMAS can perform well in both grid and slits configuration, it manages to resolve both setups.

The grid mode solves the source distribution to two pixels possible source location which approximately corresponds to 660 cm^2. Slit mode is dependent on a specific distance input, the pixel row of the FBF $\upsilon = 5$ with highest response is chosen and that specific pixel row is then re-beamformed using the same settings as before. Since slit mode resolution is reduced to just 1 dimension, the resolution per slit is approximately 22 cm, which translates to a possible source localization of 66 cm or three pixels. Comparing grid and slit mode, placed the source at approximately the same X-coordinate location with some small offset. This difference also highlight that for sound source localization purposes using grid based methods, a user can be fooled by the resolution. Managing to solve a source for 1 pixel, does not necessarily imply that the location is spot on and 100% exact, instead the truth is rather that the sound source is probably located in the vicinity of this location, and as number of pixels in a solved source area shrinks so does the possible true source area.

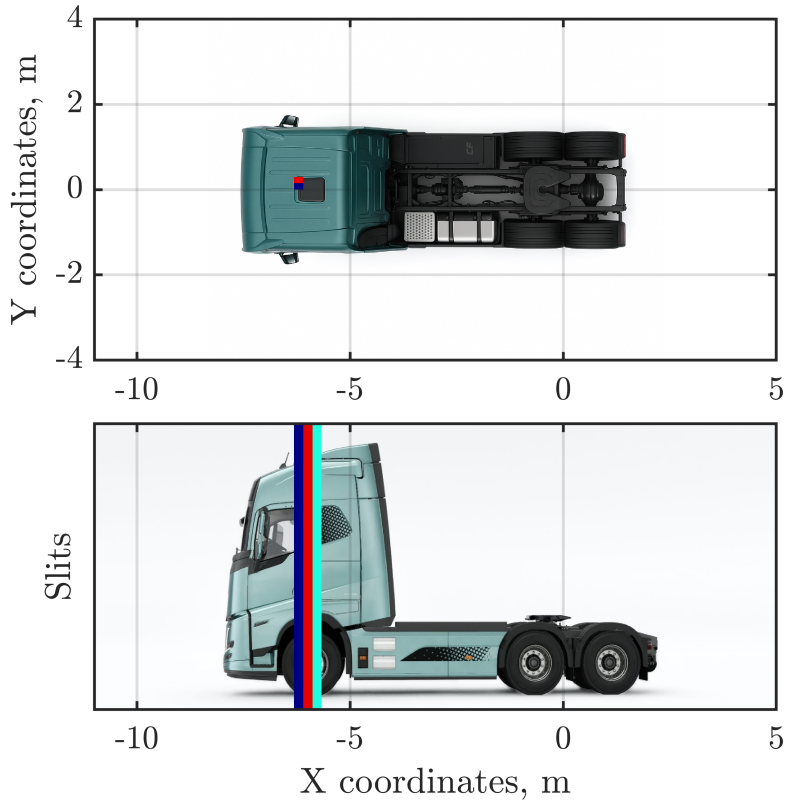


Figure 8.13: DAMAS in grid and slit mode.

Summary

The real data test was a success as every implemented technique except FBF with negative v was able to find a source position that aligns consistently comparing individual techniques. Volvo NVH engineers confirmed this seems to be the correct position of the component which emitted the tonal noise.

The setup and analysis performed here is similar to how the software and individual scripts are aimed to be used in a real-world scenario. The main purpose, localizing tonal components in a sound field with timbre similar to motor vehicles has been achieved using the provided test measurement data.

An important note relating to the normalized dB scale, seen in figure 8.9, which applies to all figures. It has previously been stated that for sound localization purposes a linear scale representation could be beneficial but that it was not possible to achieve a globalized, normalized linear scale which makes comparison between individual techniques possible. As almost all techniques manages to correctly localize a target, it is rather a question of scaling when applying the normalized dB scale. In this setup one may be inclined to draw the conclusion that DS performs worse since its array pattern obviously has more peaks in the scale-relevant range, but just changing the normalized dB scale to 0 dB to -3 dB instead of 0 dB to -6 dB would significantly reduce visible array pattern lobes. The opposite argument could be applied for MVDR, increasing the normalized dB range to -9 dB would impact

the visible parts of the array pattern, showing more coloured pixels. Because of this, it is important to draw the correct conclusions. The DS beamformer exhibits more peaks within the 0 dB to −6 dB normalized dB scale compared to the other beamformer techniques. Those techniques, which correctly identified the source location, achieves better side-lobe suppression and narrower mainlobe beamwidth resulting in less smearing and artefacts in the same 0 dB to −6 dB normalized dB scale. However, all beamforming techniques except FBF with negative v, managed to locate the sources true position. Different dB-scale range would affect how much of the array pattern is visible in the figures and it is a question of scaling when interpreting these results.

9
Discussion

This chapter presents a discussion regarding the results and findings related to each beamformer technique and DAMAS. Interpretation of individual technique results and comparison. Some important factors regarding the behaviour of the beamformer techniques are discussed, which can be translated to beamformers in a more general sense, applicable beyond the limitations of this master's thesis. Additionally, some environmental behaviour regarding the acoustic camera software performance in the semi-anechoic chamber is discussed.

The simulated beamformer validation metrics for DOA estimation showed a similar valid frequency range for all beamformers techniques where on-axis frequency range was estimated to be 40 Hz to 1600 Hz and off-axis 40 Hz to 1250 Hz, see figures 8.1, 8.2 and table 8.1. At low frequencies and in noisy conditions the DS beamformer performance is reduced compared to MVDR and FBF which is primarily due to its noise sensitivity, primarily caused by the large mainlobe width it expresses at low frequencies. This was the the primary reason continuing specifically with MVDR as the next implemented technique after DS. The validation metrics results can be compared to the pulse train measurements based DOA estimation results (figure 8.5), where cancellation effects due to the floor reflection impacts the valid frequency range, examples of reflection behaviour versus frequency range can be seen in figure 8.6.

The floor reflection obviously limits the beamformers ability to correctly source locate at certain frequencies which impacts their reliable performance range. But is it a feature or a drawback? The main purpose of this semi-anechoic chamber measurement setup, with a dynamometer placed in the central part of the room, is to simulate a vehicles road movement in an indoor controlled environment. The rig and the ULAs are designed to perform pass-by measurements, which aims to further emulate a scenario of a truck passing-by a certain location. Thus the floor reflection is part of and similar to what a listener would experience, listening to the vehicle from the side of a road. In a real world pass-by scenario the truck would move, not staying idle, and as seen in the pulse train combined with floor reflection figures 8.6, the destructive interference dips frequency location changes with distance between source and microphone. This won't happen in a semi-anechoic chamber with the vehicle stationary on a dynamometer, but a feature of this non-moving scenario is the possibility to estimate the vehicles radiation characteristics in a road-like controlled environment. This behaviour is also captured by our beamformers and their reduced performance at the destructive interference dips because of source radiation characteristics. However this is still more of a drawback than a feature since, in the end, it impacts and reduces the reliable frequency range of the beamformers.

A limiting factor in the beamformers performance in this specific scenario is spatial resolu-

tion. Table 4.1 presents spatial resolution estimates for all 1/3-octave band center frequencies. This applies to all beamforming techniques and demonstrates their inability to separate 2 sources spaced at shorter, or similar distances as their wavelength λ. This is expected, it is a general limitation for beamformers. A main focus of this master's thesis has been source localization of distinct tonal components in a vehicle-like timbre. But, as the frequency range of interest extends down to 50 Hz where λ is 6.87 m, spatial resolution becomes a problem due to large λ and beamformer mainlobe widths. This was the primary reason for implementing the post-processing algorithm DAMAS as the last technique. It has the ability to clean SRPs and pixel maps from smearing effects and artifacts. DAMAS slightly improves the overall spatial resolution of the acoustic camera software. At 50 Hz it manages to resolve sub-λ distanced sources by 40 cm, which compared to the λ size, is not much. This inability to separate lower frequencies sources is one of the biggest flaws of this acoustic camera software. The primary contribution of DAMAS lies in its ability to clean smeared SRPs and pixel maps, providing improved interpretability in complex sound field scenarios. Its usefulness is most apparent when coupled with FBF and its order factor υ, which introduces tunability to SRPs through side-lobe suppression and control over the mainlobe width. This provides a valuable extra degree of control to the input SRPs and pixel maps into DAMAS. This also underscores that DAMAS is fundamentally different than the beamformer techniques, as it is entirely dependent on the input from the beamformers to operate. This makes comparison between the implemented beamformers and DAMAS awkward, as the DAMAS output can never be separated from the performance of the beamformer input.

All beamformer techniques that has been implemented and evaluated, except FBF with negative υ, was able to find real data measured source as seen in the vehicle measurements results, see subsection 8.3. There are distinct differences in the beamformers ability to suppress smearing and side-lobe artefacts, but all techniques still managed to correctly localize the source. Using the acoustic camera software in a real-world scenario, where a non-localized tonal component was present in the sound field, highlighted the importance and strength of using multiple techniques combined to achieve both precision and confidence in the results. The valid frequency range estimates based on the pulse train measurements (see figure 8.5) also visualizes that there are distinct possibilities that one specific beamformer may not be able to localize a source at a certain frequency and location, while another beamformer technique will. Therefore it is also a question of robustness of the combined acoustic camera software, using multiple available beamformer techniques when evaluating specific sound field in order to achieve a valid and accurate result.

A recurring observation throughout the thesis has been the importance of accurately computing the timing delays τ and creating precise steering vectors. Even small errors in calculation of τ can introduce offsets in the array response, especially at higher frequencies. A probable example of this could be seen comparing on-axis validation metric DOA estimations in figures 8.1, 8.2 or table 8.1, with the valid frequency range of the pulse train measurements, position 2 high SNR scenario, figure 8.6. The FBF manages to correctly estimate a source position, unphased by the floor reflection up until 2 kHz. This is higher than what the validation metrics results imply is the upper frequency range, its valid frequency range is limited to 1600 Hz. The validation metrics offsets is larger, at ± 5 cm, than what the construction of the microphone stands in the semi-anechoic chamber would allow, figure 3.6 is a picture

from the pulse train measurements where the microphone stand construction can be seen. This deviancy could therefore be an effect of the real-life smaller microphone positions offsets which directly impacts τ and the precision of the steering vectors. The FBF has $\upsilon = 10$ which is probably what makes it power through the floor reflection destructive interference. This increased frequency range in measurements compared to simulated validation metrics estimations cannot be verified at any other positions or further examined with the available data, since the floor reflections cannot be extracted nor separated from the offsets in the real-life measurements.

Examining microphone sensitivity mismatches showed no variability within the valid frequency range, even with a ±20 % random gain offset. This offset is considered high as even larger deviations should noticeably affect the structure of the CSM. Since the CSM represents a measurement of similarity comparing microphone signals, substantial sensitivity mismatches are expected to degrade the SRP. Figure 5.1 shows the CSM for a uniform gain configuration, introducing larger gain offsets would degrade this structure even more which would result in a degraded SRP. The validation results indicate that a ±20 % gain offset is still low enough to avoid degrading the SRP to the point where the source cannot be located. A deviation as small as ±5 % would already prompt Volvo NVH engineers to recalibrate the system.

From a user perspective, each technique has its individual strengths. DS is simple to understand, computationally efficient and robust to irregularities in the measurement setup, making it suitable for quick and easy checks or high SNR environments. MVDR and FBF offers improved directionality and narrower mainlobes in worse SNR conditions. Both techniques also allow for frequency-specific beamforming, enabling the isolation of individual tonal components. This is a valuable feature for source localization purposes, where the presence of several peaks in an SRP or pixel map can confuse the user as to which peak corresponds to a specific frequency target source. FBF when combined with DAMAS further enhances the full acoustic camera software with extended source separation capabilities. A key advantage of this near-field beamforming setup is its ability to enable 2D source localization through spherical wave propagation at the location of the ULAs, even though it is inherent to their design that they have only a one-dimensional microphone distribution. This further extends the acoustic camera software's source localization abilities and examples of this can be seen in the pulse train localization figure 8.7 and in subsection8.3 depicting the real vehicle measurements.

One of the most important insights from this work is that its preferable to analyse beamforming technique results side-by-side. The full strength of the system lies in combining the techniques. Starting with MVDR for robust and spatial selective source localization, tune the output with FBF to get a clearer SRP and apply DAMAS where needed for improved source separation and reduced smearing. This analysis approach can help to achieve both reliable interpretability and localization accuracy.

10
Conclusion

This master's thesis aimed to investigate and evaluate the potential use of SULAs as part of an indoor pass-by measurement setup in a semi-anechoic environment, functioning as acoustic cameras. Three beamforming techniques has been implemented, tested and compared to asses their practical performance in both simulated and real acoustic conditions. The beamformers are DS, MVDR and FBF. Additionally, a fourth post-processing algorithm DAMAS was introduced as the last implemented technique part of the complete acoustic camera software. Each technique was evaluated using custom-designed validation metrics involving noise, array microphone misplacements and sensitivity variations to understand their robustness, spatial resolution and acoustic source localization ability.

All techniques were able to handle real measurement data recorded in the semi-anechoic chamber, although their individual performance varied depending on measurement conditions. DS offers robustness to microphone placement differences and minimal complexity but struggles in noisy environments. MVDR and FBF techniques both relies on a CSM in their implementations and offers improved interference and noise suppression capabilities at the cost of greater sensitivity to microphone placement mismatches and increased computational load. FBF introduced a tuneable order factor υ which affects the SRPs side-lobe suppression and mainlobe width. When combined with DAMAS, it enables improved spatial resolution compared to the individual beamformer techniques.

Beamforming is a broad and ever-evolving field and each technique and algorithm studied in this thesis work has potential for further development. One addition to this specific setup would be to design a model which accounts for the interference of the sound field introduced by the floor reflection. This may not, however, be straightforward. The results provided in this work are based on simulated monopole sources and measurements were conducted with loudspeakers designed to approximate monople radiation. Real acoustic sources are not perfect monopoles and their placement relative to the floor introduces complex behaviour that are difficult to model accurately.

The large spatial extent of the SULAs in relationship to the possible source area ensures the frequency range of interest, 50 Hz to 2 kHz, is within the near-field range of the array, characterized by spherical wave propagation. Despite the array being in 1D in physical configuration, these SULAs can support 2D localization due to the spherical curvature of the wavefronts.

Finally, one of the most critical observations is the steering vectors formulation and the time delay τ computation, which plays a critical role in beamformer accuracy. Even minor errors

in these calculation can significantly degrade localization performance especially in higher frequency ranges.

Key Insights

This subsection provides key-insights gathered throughout this master's thesis work. The insight are gathered and summarized in the bulleted list below.

- The DS beamformer is robust to array mismatches but performs poorly in low SNR environments.
- The MVDR beamformer provides sharper mainlobes and improved noise rejection but depends on accurate CSM estimation.
- The FBF introduces tunable trade-off using the order factor v.
- DAMAS enhances spatial resolution capabilities and can resolve complex source distributions.
- Sparse ULAs in near-field conditions enables 2D source localization, even from a 1D array geometry.
- Precise time delay and steering vector computation is essential across all beamforming techniques and directly impacts resolution and robustness.
- The singular floor reflection in the semi-anechoic chamber hampers the source localization capability of the acoustic camera software.
- The combination of multiple techniques is key for achieving accurate, interpretable and reliable results.

Bibliography

[1] *Uniform provisions concerning the approval of: Motor vehicles having at least four wheels with regard to their sound emissions*, UNECE 51-03, Geneva, Switzerland: United Nations Economic Commission for Europe, 2019.

[2] *Measurement of Noise Emitted by Accelerating Road Vehicles – Engineering Method – Part 1: M and N Categories*, ISO 362-1, Geneva, Switzerland: International Organization for Standardization, 2022.

[3] *Measurement of Noise Emitted by Accelerating Road Vehicles – Engineering Method – Part 3: Indoor Testing of M and N Categories*, ISO 362-3, Geneva, Switzerland: International Organization for Standardization, 2022.

[4] M. M. Erić, "Some research challenges of acoustic camera," in *2011 19thTelecommunications Forum (TELFOR) Proceedings of Papers*, 2011, pp. 1036–1039. DOI: 10.1109/TELFOR.2011.6143725.

[5] J. Grythe. "Norsonic acoustic camera - filming sound leakages in highly reflective office environment." (2016), [Online]. Available: https://norsonic.se/wp-content/uploads/2023/09/Norsonic-acoustic-camera-case-study-office-leakage.pdf (visited on 04/28/2025).

[6] T. D. Rossing, Ed., *Handbook of Acoustics*. Springer, 2014.

[7] A. V. Oppenheim, R. W. Schafer, and J. R. Buck, *Discrete-Time Signal Processing*, 2nd. Upper Saddle River, NJ: Prentice Hall, 1999, ISBN: 978-0-13-754920-7.

[8] J. Benesty, J. Chen, and Y. Huang, *Microphone Array Signal Processing* (Springer Topics in Signal Processing). Springer, 2008, vol. 1, ISBN: 978-3-540-78611-5. DOI: 10.1007/978-3-540-78612-2. [Online]. Available: https://link.springer.com/book/10.1007/978-3-540-78612-2.

[9] P. Delos, B. Broughton, and J. Kraft, "Phased array antenna patterns—part 2: Grating lobes and beam squint," *Analog Dialogue*, vol. 54, no. 2, Jun. 2020, Accessed: 2025-08-06. [Online]. Available: https://www.analog.com/en/analog-dialogue/articles/phased-array-antenna-patterns-part-2.html.

[10] A. Patwari, "Sparse linear antenna arrays: A review," in Sep. 2021, ISBN: 978-1-83968-828-7. DOI: 10.5772/intechopen.99444.

[11] S. Sun, R. Li, C. Han, X. Liu, L. Xue, and M. Tao, "How to differentiate between near field and far field: Revisiting the rayleigh distance," *IEEE Communications Magazine*, vol. 63, no. 1, pp. 22–28, 2025, ISSN: 0163-6804. DOI: 10.1109/MCOM.001.2400007. [Online]. Available: https://libkey.io/10.1109/MCOM.001.2400007.

[12] C. A. Balanis, *Antenna theory: analysis and design*. Wiley-Blackwell, 2012.

[13] J.-W. Choi and Y.-H. Kim, "Spherical beam-forming and music methods for the estimation of location and strength of spherical sound sources," *Mechanical Systems and Signal Processing*, 1995.

[14] G. C. Carter, "Coherence and time delay estimation," *IEEE*, 1987.
[15] J. Allen and D. Berkley, "Image method for efficiently simulating small-room acoustics," *The Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, Apr. 1979. DOI: 10.1121/1.382599.
[16] *Specification of test tracks for measuring noise emitted by road vehicles and their tyres*, ISO 10844, Geneva, Switzerland: International Organization for Standardization, 2021.
[17] D. H. Johnson and D. E. Dudgeon, *Array Signal Processing: Concepts and Techniques*. Prentice Hall, 1993.
[18] V. Valimaki and T. Laakso, "Principles of fractional delay filters," in *2000 IEEE International Conference on Acoustics, Speech, and Signal Processing. Proceedings (Cat. No.00CH37100)*, vol. 6, 2000, 3870–3873 vol.6. DOI: 10.1109/ICASSP.2000.860248.
[19] D. Fernandez Comesaña, K. Holland, and E. Fernandez-Grande, "Spatial resolution limits for the localization of noise sources using direct sound mapping," *Journal of Sound and Vibration*, vol. 375, pp. 53–62, 2016. DOI: 10.1016/j.jsv.2016.04.010.
[20] J. Capon, "High-resolution frequency-wavenumber spectrum analysis," *Proceedings of the IEEE*, vol. 57, no. 8, pp. 1408–1418, 1969. DOI: 10.1109/PROC.1969.7278.
[21] M. Lichtman, *Beamforming & doa*, [online] https://pysdr.org/content/doa.html, Accessed: 30-06-2025.
[22] N. Ma and J. T. Goh, "Efficient method to determine diagonal loading value," in *2003 IEEE International Conference on Acoustics, Speech, and Signal Processing, 2003. Proceedings. (ICASSP '03).*, vol. 5, 2003, pp. V–341. DOI: 10.1109/ICASSP.2003.1199948.
[23] Y. Xiao, J. Yin, H. Qi, H. Yin, and G. Hua, "Mvdr algorithm based on estimated diagonal loading for beamforming," *Mathematical Problems in Engineering*, vol. 2017, no. 1, p. 7 904 356, 2017. DOI: https://doi.org/10.1155/2017/7904356. eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1155/2017/7904356. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1155/2017/7904356.
[24] J. L. Subbiah and R. Sukanesh, "A novel robust adaptive beamformer using diagonal loading," *International Journal of Computer Applications*, vol. 19, Apr. 2011. DOI: 10.5120/2355-3079.
[25] C. Wu, Y. Guo, Y. Na, X. Wang, Q. Fu, and Y. Yan, "Robust beamforming using beam-to-reference weighting diagonal loading and bayesian framework," *Electronics Letters*, vol. 51, no. 22, pp. 1772–1774, 2015. DOI: https://doi.org/10.1049/el.2015.2859. eprint: https://ietresearch.onlinelibrary.wiley.com/doi/pdf/10.1049/el.2015.2859. [Online]. Available: https://ietresearch.onlinelibrary.wiley.com/doi/abs/10.1049/el.2015.2859.
[26] L. Ehrenberg, S. Gannot, A. Leshem, and E. Zehavi, "Sensitivity analysis of mvdr and mpdr beamformers," in *2010 IEEE 26-th Convention of Electrical and Electronics Engineers in Israel*, 2010, pp. 000 416–000 420. DOI: 10.1109/EEEI.2010.5662190.
[27] M. Nezafat, M. Kaveh, and H. Tsuji, "A robust mvdr spectrum estimation technique," in *2007 IEEE Workshop on Statistical Signal Processing (SSP)*, Accessed August 7, 2025, Madison, WI, USA: IEEE, 2007, pp. 551–555. DOI: 10.1109/SSP.2007.4301284. [Online]. Available: https://doi.org/10.1109/SSP.2007.4301284.
[28] Dougherty, "Functional beamforming," in *Proc. 5th Berlin Beamforming Conference (BeBeC)*, Accessed: 11-07-2025, GFaI, Gesellschaft zu Förderung angewandter Infor-

matik e.V., Berlin, Feb. 2014, ISBN: 978-3-942709-12-5. [Online]. Available: http://bebec.eu/Downloads/BeBeC2014/Papers/BeBeC-2014-01.pdf.

[29] B.-H. Wu and N.-H. Chang, "Application the deconvolution deblurring to nearfield beamforming for acoustic array," in *2024 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2024, pp. 1–5. DOI: 10.1109/ISPACS62486.2024.10868214.

[30] T. Brooks and W. Humphreys, "A deconvolution approach for the mapping of acoustic sources (damas) determined from phased microphone arrays," vol. 2, May 2004. DOI: 10.2514/6.2004-2954.

[31] Volvo Trucks, *Truck builder*, [online] https://www.volvotrucks.com/en-en/trucks/information-and-tools/truck-builder.html, Accessed: 2025-07-31, 2025.

[32] OpenAI ChatGPT (GPT-5), *From-above truck image generated using chatgpt*, [online] https://chat.openai.com/, Generated and accessed on 2025-07-31, 2025.

DEPARTMENT OF SOME SUBJECT OR TECHNOLOGY
CHALMERS UNIVERSITY OF TECHNOLOGY
Gothenburg, Sweden
www.chalmers.se

CHALMERS
UNIVERSITY OF TECHNOLOGY