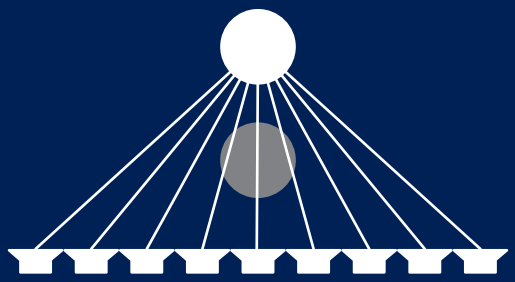
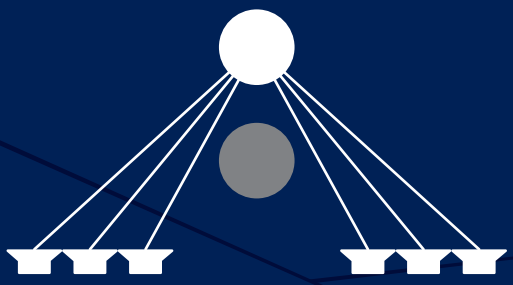
INVESTIGATION OF PERSONAL SOUND ZONES BY USE OF SUBARRAY DIVISION

SIGNAL PROCESSING AND ACOUSTICS





Master's Thesis | June 4, 2020

Jonas Bredgaard Köhne

Mikkel Solhaug Seindal

Supervised by Christian Sejer Pedersen

BREAKING NEW GROUND

AALBORG UNIVERSITY

AALBORG UNIVERSITY
STUDENT REPORT

Aalborg University
Department of Electronic systems
Signal Processing and Acoustics
Fredrik Bajers Vej 7
9220 Aalborg Øst
www.es.aau.dk

Title:

Investigation of Personal Sound Zones by Use of Subarray Division

Theme:

Signal Processing and Acoustics

Period of the project:

February 2020 - June 2020

Project group:

20gr1076

Authors:

Jonas Bredgaard Köhne
Mikkel Solhaug Seindal

Supervisor:

Christian Sejer Pedersen

Pages: 121
Appendix: A - F
Ended: 04-06-2020

Abstract:

The project investigates a method of dividing a loudspeaker array into smaller subarrays in order to create higher contrast between to personal sound zones with sub-optimal positions.
The weighted pressure matching algorithm is used with a line array to create personal sound zones. Simulation tests showed that an increase in contrast was achieved when turning off loudspeakers. Ray space edge subdivision (RSES), a method, using the ray space transform, to determine which of the loudspeakers that should be turned off was made.
The method was tested in a measurement room, with a line array consisting of 39 loudspeakers. Two different sub-optimal setups was tested, setup 1) with the sound zones placed with centers in (0,0.5) and (0,1.5), and setup 2) with centers in (0.25,1.5) and (-0.25,1.5). Both relative to the center of the loudspeaker array. The frequency region of interest is 315-4000 Hz, and is based on the dimensions of the loudspeaker array.
The results show that an increase in contrast is achieved when using the proposed method, with an increase in the mean square error. At higher frequencies an increase of up to 12 dB is achieved with the RSES-model, No improvement is seen at the lower frequencies. However the results are very dependent on the specific setup. An informal listening test and spectrograms show that the increase in MSE, is due to small changes in amplitude.

The content of the paper is freely available, but publication (with source reference) may only take place in agreement with the authors.

Preface

This thesis is conducted by Jonas Bredgaard Köhne and Mikkel Solhaug Seindal (group 20gr1076), Signal Processing & Acoustics at Aalborg University. The group would like to thank Engineering Assistant Claus Vestergaard Skipper for technical assistance.

Reading Instructions

This report is divided into chapters numbered by the order in which they appear. Sections and subsections are numbered following the same methodology, but subsubsections appear without numbers.

Figures, tables, equations and code examples are numbered by the chapter they appear in as well as the order. E.g. figure 5 in chapter 4 will be numbered as 4.5. Appendices are found in the end of the report and will be numbered with letters, starting from **A**, which will be referred to throughout the report.

The reference method *Vancouver* is used throughout the report. A list of the references is found at the end of the report, numbered and organized by the order they are used in the report. References are indicated in the text with a number that matches the list in the back.

Relevant scripts can be found in a Dropbox folder using following link:
https://www.dropbox.com/sh/2ggo8jza86w5o7a/AADofgZZC18bWDROs5BN-afKa?dl=0

A demo of the system can be found by using following link:
https://www.youtube.com/channel/UCreJ9ZWh51PoUBAXMNNcXLQ?view_as=subscriber

Jonas Bredgaard Köhne
<jkahne15@student.aau.dk>
<jonas-kohne@hotmail.com>

Mikkel Solhaug Seindal
<mseind15@student.aau.dk>
<mikkel@seindal.net>

Contents

1 Introduction

Personal sound zones gives the ability for a listener to have their own audio zone without physical isolation or the use of headphones [1], which would isolate the individuals and inhibit social interaction. The usage of personal sound zones can be many. E.g. patients in a row of hospital beds may be watching different television stations, museum exhibits may feature associated audio tracks [2] or in a living room where a TV sound system (e.g. a soundbar) boost the frequencies and audio level in a certain area where hearing impaired listeners are seated [3]. In the last example, the goal is to minimize the radiation in all other areas than the target zone, so that the normal hearing perceive the audio without being disturbed [3]. A typical scenario as the one described is illustrated in figure 1.1.

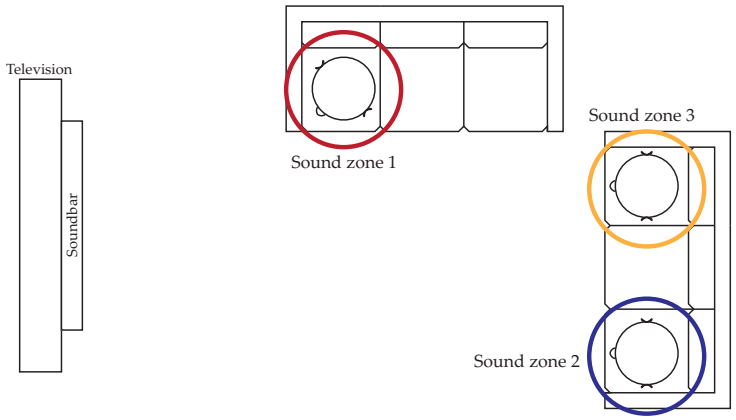


Figure 1.1: Illustration of a living room scenario with three sound zones.

The aim of the personal sound zones is to create an acoustically bright and an acoustically dark zone, with minimum leakage between the two zones [4]. The bright zone refers to an area where the sound should be reproduced with high acoustic energy, and the dark zones refers to an area where the goal is to minimize the acoustic energy [1], this approach relies on superposition [5]. Superposition assumes a linear system and can be used to describe the total pressure in a point as the sum of all the pressure contributions. If more listeners are in the same room and should have their own zone, the bright zone for one person will be a dark zone for the other. If there are Q sound zones, there will be one bright zones and $Q-1$ dark zones, for each listener, meaning that all other zones than the bright zone will be referred as dark zones [1], this is illustrated in figure 1.2.

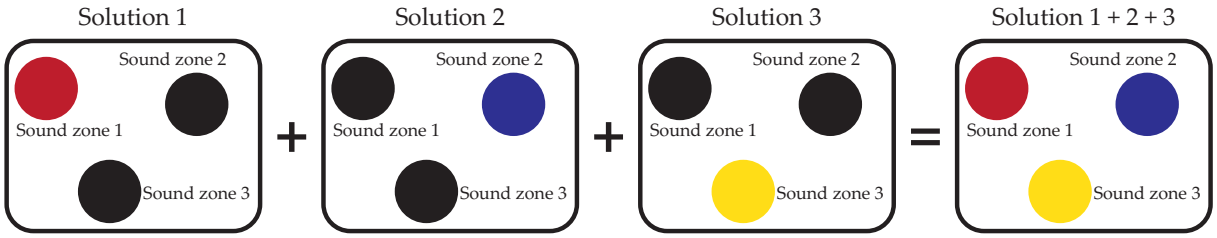


Figure 1.2: Illustration of the principle of superposition, with three different sound zones. In solution 1, sound zone 2 and 3 are seen as dark zones, solution 2 are sound 1 and 3 seen as dark zones and solution 3, 1 and 2 are seen as dark zones.

Different Types of Sound Zones

As mentioned earlier there are different applications for sound zones. Where the absolutely optimal setup for a personal sound zone would be to simulate the use of a headset, where you can move around freely, listening to your own audio, with no leakage to the other zones but without being insulated from the surroundings. This would have to be feasible for the entire audio spectrum, with no limitations to spacing of the sound zones. However this is both very complex, expensive and almost impossible to realize all of the features at once. In most cases compromises of these features can be made to reduce the complexity and expense of a practical solution, but still achieve a desired performance for specific scenarios. An example of this is e.g. a row of hospital beds, where it is preferred to work for the entire audio spectrum in order to get the best listening experience. However, because the beds are stationary there is no desire for the zones to be movable. Another application could be in a conference hall, e.g. used for global politics, where different listeners might need to listen to the speakers in a different language. As the audio signal in this scenario almost exclusively contains speech signals, the frequency range can be reduced to only be within the range of speech. In this application, the listeners are seated closer than in the hospital scenario meaning that a higher requirement for the preciseness of the sound zone would be required in order to make up for the smaller spacing of the sound zones. It is again not desired to be able to move the sound zones. Whereas in the living room scenario the sound zones are in a much higher degree preferred to be movable, since it allows for movement within the living room. However here the zones are not expected to be as close as in the former mentioned conference hall scenario.

Different Methods for Creating Sound Zones

Sound zones can be made using an array of loudspeakers, but it should be noted that different methods are more suitable for some frequency regions. In the mid-range frequencies, the beamforming technique can be used. This technique uses the knowledge about constructive and destructive interference, to direct a beam in a certain direction using an array of loudspeakers. Due to the natural directivity of the loudspeakers at higher frequencies the beamforming technique is not viable here, as overlap of the loudspeakers are required in order for it to work efficiently. For these frequencies, simple pointing a loudspeaker in a certain direction can often provide good results. Likewise there is a lower limitation of the frequencies viable

in beamforming, which is that the array should be much larger than the wavelength [6], meaning that for a setup at home the loudspeaker array would simple be too large in order for beamforming to be a viable solution at the lowest frequencies. Here active sound control can be used instead. This technique makes use of some strategically placed loudspeakers around in the room. Summed up, this means that in general three different methods are considered when trying to reproduce the sound in a desired region while trying to reduce the sound in other regions. The methods are: active control techniques which are used at low frequencies ($< 1\,\text{kHz}$), conventional beamforming is used to mid-range frequencies (1 - 3 kHz) and for high frequencies ($> 3\,\text{kHz}$), directive loudspeaker is used [7]. This is conceptually illustrated in figure 1.3. It should be noted that the frequency range described can variate, dependent on the loudspeaker array.

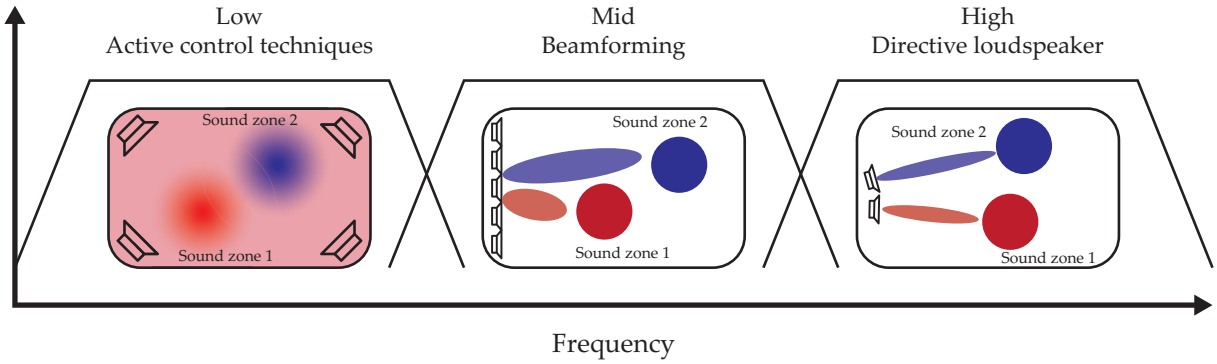


Figure 1.3: Illustration of the different methods of creating sound zones at different frequencies.

If the application should be an ”out of the box” solution in form of a soundbar to place under the TV, the setting up procedure without a technician to help, will be preferable. Thus, no measurements of the room should be made. Here the low frequencies will often require more knowledge about the room due to the physical properties of the frequencies, such as wave length and the time it takes the frequencies to die out. Therefore, beamforming at mid-range frequencies will be preferable, as it allows an array of loudspeakers to steer an audio stream in a specific direction. However the performance of the beamforming is still sensitive to the room response due to the reflections in the room, thus the robustness of a proposed system should be tested for different scenarios, and how sensitive it is to changes in the transfer function of the room. A scenario could be increased audio level for hearing impaired listeners in the speech frequencies region, which contains mid-range frequencies, or if the same movies is available in two different language, two different persons could see the movie in their preferred language. In this scenario the low and the high frequencies are not important to separate as in a movie these frequencies mostly contain the background music and the sound effects which is the same for both listeners.

Based on a living room situation, where a soundbar is the loudspeaker array (linear array), some different challenges such as: the amount of loudspeakers accessible, the room and also the seating positions of the listeners, can affect the creation of the sound zones. The amount of loudspeakers is important in order to achieve good performance as more loudspeakers allows for a more precise reproduction of a desired signal in a certain area, which then allows for more and smaller sound zones [1]. E.g. tests described in [1] tells that using an array of three loudspeakers have showed a contrast between the bright and dark zone of 10 dB, compared to an array of nine loudspeakers which gives more that 19 dB contrast in an anechoic chamber. However when

this setup is moved outside of the anechoic chamber, a well known challenge is the reverberation of the room. This can cause an impaired performance when trying to reproduce a sound field [1].

Another challenge is that with sub-optimal position of the individual sound zones, there will in some cases be sound zones shadowing for each other for certain loudspeakers in the loudspeaker array. Two examples of this are illustrated in figure 1.4. The examples are based on the living room scenario with a line array. If the loudspeakers where distributed differently, this might not be an issue or introduce some other challenges.

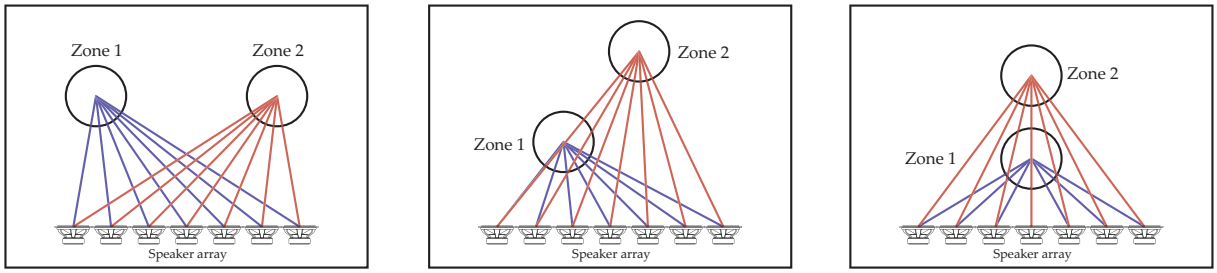


Figure 1.4: Illustration of three different positions of a two different sound zones. In two of the examples one of the sound zones are in the shadow of the other. The blue lines show the path from the loudspeaker to the first zone and the red lines show the path to the second zone. Where shadowing occur if the lines are crossing through one of the sound zones.

As it can be seen illustrated in figure 1.4 the audio stream from the loudspeaker in the direct path to the zone, the leakage between the zones will be expected to be higher if the line(s) are crossing another zone. An idea to solve this problem is simply to turn off the loudspeakers where one zone are blocking the direct path to the other. This concept is illustrated in figure 1.5.

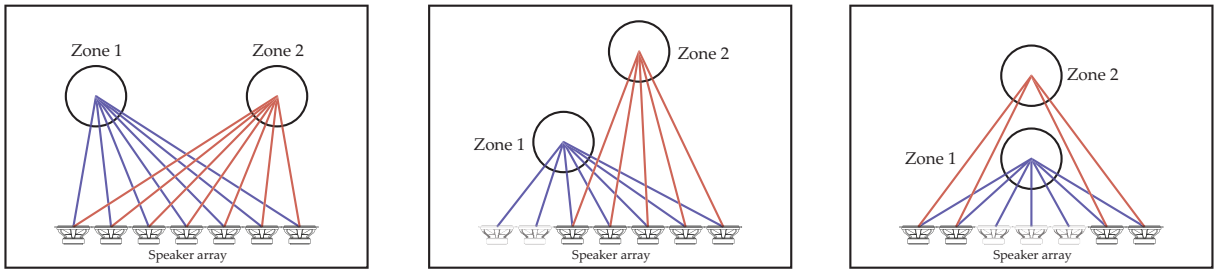


Figure 1.5: Illustration of three different positions of a two different sound zones showing the principal of turning off loudspeakers where the direct paths to a zone are blocked by another zone.

The light grey loudspeakers are the ones that should be turned off when having Zone 2 as bright and Zone 1 as dark. However, as it can bee seen in the figure, when having Zone 1 as bright and Zone 2 as dark, no direct path from the loudspeaker array to the zone are blocked, thus it is assumed that there is no need to turn off these loudspeakers in this configuration.

In this project a line array will be used to recreate the living room scenario described above, meaning that the focus is beamforming and the mid-range frequencies. A study of different state of the art algorithms used for personal sound zones, and an investigation in order to determine their performance in sub-optimal setups (illustrated in figure 1.4) will be made. A proposed

method of how to separate the array into subarrays will be made in an attempt to reduce the leakage between the sound zones. Leading up to the problem statement.

1.1 Problem Statement

Can the performance of beamforming be improved in sub-optimal positions by dividing the line-array into smaller subarrays and strategically removing specific loudspeakers?

2 | Multizone Sound Control

In this chapter, the concept of personal sound zones and the general principles when using a linear loudspeaker array will be described. Then two fundamental sound zone control techniques and their properties will be described. Next, a state of the art algorithm will be reviewed and simulated to investigate its performance. Here, a summary will be made of the parameters that can have influence on the performance and robustness of the algorithm when trying to create personal sound zones. Finally different methods of how to evaluate the results is proposed and described.

2.1 Signal Model and the Delay and Sum Beamformer

Before discussing the properties of the sound zones, a generalized signal model will first be introduced. The basics of controlling the sound field can be achieved by using an array of loudspeakers [8]. When the loudspeakers are all fed the same original signal with different filtering, the system are able to focus the acoustic pressure in certain positions - this is called beamforming [1]. The simplest one being the delay and sum (DS) beamformer, which just introduces a delay on each of the loudspeakers [6]. Beamforming uses constructive and destructive interference in order to change the amplitudes of a signal based on the direction [9]. A system like this can be seen illustrated in figure 2.1.

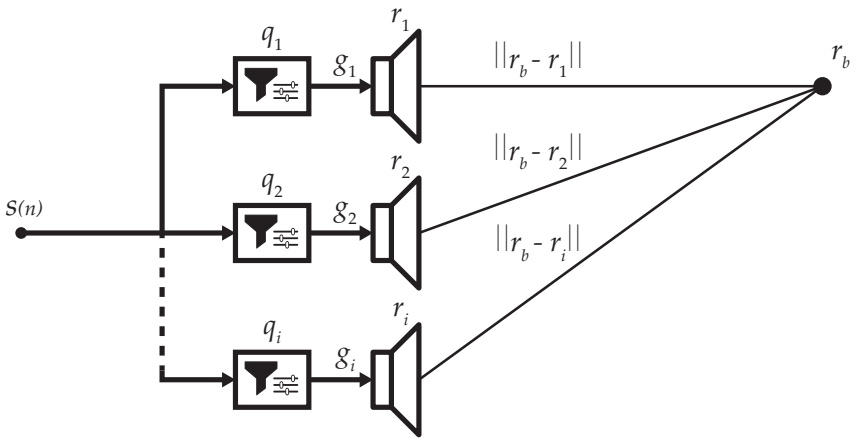


Figure 2.1: Diagram for a simple beamformer.

The signal played from the loudspeakers can be described with the following equation:

$$g_i(n) = s(n) * q_i(n) \tag{2.1}$$

Where $s(n)$ is the source signal fed to all of the loudspeakers, q_i is the individual filter for each loudspeaker, and g_i describes the signal played from the loudspeakers, with $i = 1, 2, ..., L$. Based on this, the pressure in the point, $P_{r_b}(n)$, can be described as a convolution of the signal played

from the loudspeakers and the acoustic transfer function of the room, h. This can be written as:

$$P_{r_b}(n) = \sum_{i=1}^{L} g_i(n) * h_i(n) \tag{2.2}$$

This transfer function can be simply estimated with a free field assumption as a delay, based on the distance between a loudspeaker and a point, and a reduced pressure based on the distance square law. This can be written as:

$$h_i = \frac{\delta\left(n - \frac{\|r_b - r_i\|}{c}\right)}{4\pi \|r_b - r_i\|} \tag{2.3}$$

Where $r_b \in \mathbb{R}^3$ is the position of the target point and $r_i \in \mathbb{R}^3$ is the position of the loudspeaker. It is known that a convolution in the time domain is a multiplication in the frequency domain. In order to reduce the computational complexity, the signal model is transformed into the frequency domain using the discrete Fourier transform (DFT):

$$\hat{P}_{r_b}(\omega) = \sum_{i=1}^{L} g_i(\omega) \frac{\mathrm{e}^{-j\omega\frac{\|r_b - r_i\|}{c}}}{4\pi \|r_b - r_i\|} \tag{2.4}$$

Two examples of the DS-beamformer can be seen in figure 2.2, one is aiming towards a point and the other aiming in a direction as well as an example without any beamforming. It uses the structure of the diagram showed in figure 2.1. Where the delay is calculated in order to maximizes the amplitude in a single direction. However this algorithm does not make any effort to minimize the amplitude in other directions.

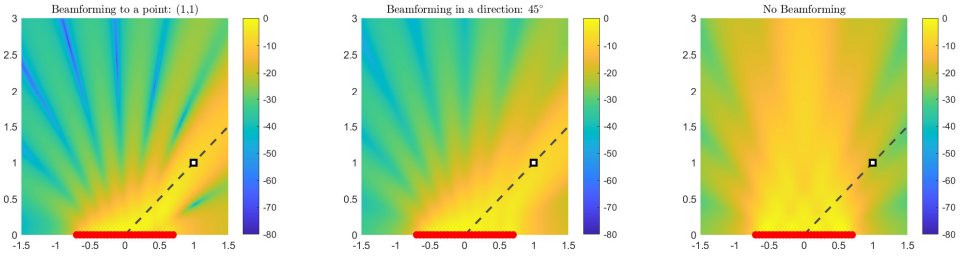


Figure 2.2: illustration of DS-beamforming at 1 kHz, with the first figure beamforming to a point $(1, 1)$, the second beamforming in a direction at 45° and the third without beamforming. The array consists of 29 point sources placed with a distance of 0.05 m, and center in $(0, 0)$).

As it can be seen in figure 2.2 both beamforming to a point and in the direction of the point, provide an increased amplitude in that direction compared to the one without beamforming. Here it can be seen that beamforming to a point, provides a more narrow beam, compared to beamforming in a direction and that natural directivity occurs when no filtering is done. However a lot of unwanted sound is radiated in all directions and would, in an actual room, create a lot of reflections. By using optimal filters, the radiation in the directions that are not of interest can be reduced. This is shown later where DS-beamforming will be compared to more advanced beamforming techniques, by their ability to create a contrast between two control points.

2.2 Techniques for Multizone Sound Control

The method of making multizone sound control is to formulate an optimization problem which describes a separation of the sound field of a physical space into, Q, smaller subareas while still being restricted to the same physical space, this is called sound zones [1]. A block diagram of a sound zone system can be seen in figure 2.3. The basic idea is that different filters are calculated and applied to the signal for each loudspeaker in order to make a sound zone separation. The filters are calculated with an optimization algorithm and are based on the acoustic transfer functions. The filtered signals are then played in the loudspeaker array and the sound field in each zone can then be evaluated to see the system performance.

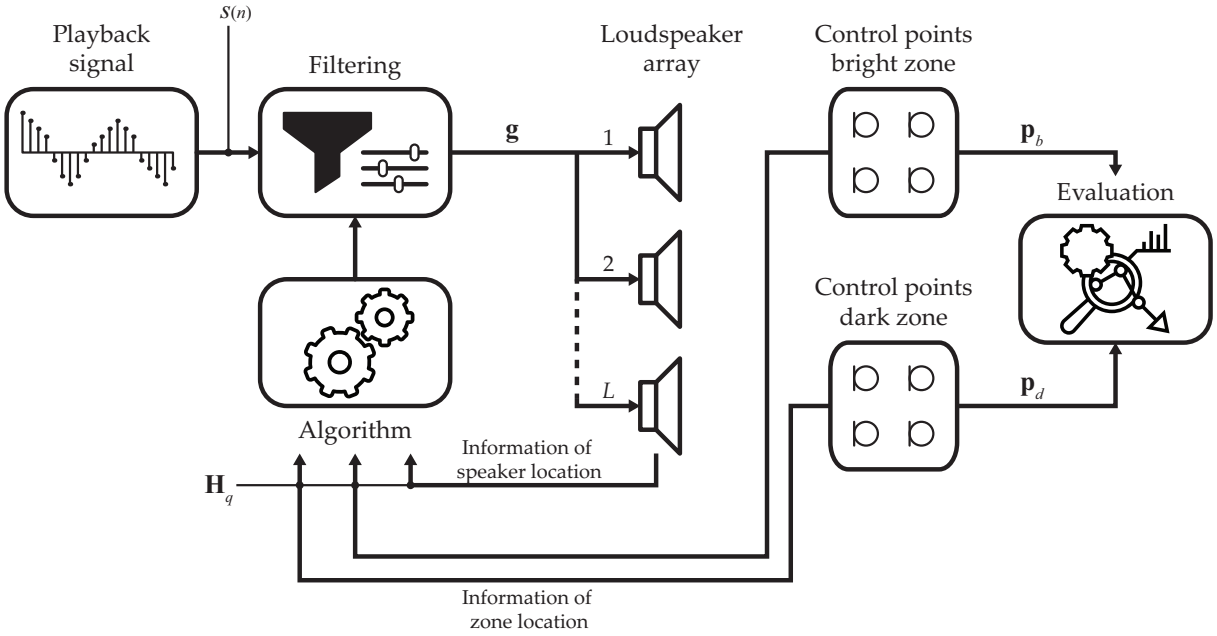


Figure 2.3: Block diagram of the signal path in the system for creating sound zones with a loudspeaker array.

The optimization problems are dependent on two types of control points. The ones in the bright zone where the goal is to produce a desired sound field and maximizing the acoustic energy and the ones in the dark zone is where the goal is to minimize the acoustic energy. The control points consist of M measuring microphones in each of the Q sound zones, for a total of QM control points. This is used to reduce the amount of leakage between the bright zone and the dark zones. The measured sound pressure for each of the control points are represented as a vector: $\mathbf{p}_q = [p(x_q, 1, \omega), ..., p(x_q, M, \omega]$ with $x_q = \mathbf{h}_q\mathbf{g}$ and given by:

$$\mathbf{p}_q = \mathbf{H}_q \cdot \mathbf{g} \tag{2.5}$$

Where $\mathbf{g} = [g_1(s, \omega), ..., g_L(s, \omega)]$, is a vector of the loudspeakers driving signals for a given frequency, $\omega = 2\pi f$. $\mathbf{H}_q$ is a matrix of the acoustic transfer functions between the loudspeakers in the array and the M control points in sound zone $q \in 1, 2, ..., Q$ [1]. Most state of the art algorithms used for sound zone control, is based on sound control techniques that can be broadly classified in two different categories: **acoustic contrast control** (ACC) and **pressure matching** (PM) [1]. These techniques will be described in the following subsections.

2.2.1 Acoustic Contrast Control

The acoustic contrast control is an optimization problem that aims to maximize the ratio between the average potential acoustic energy in the bright zone and the dark zone. The average potential energy can for the bright and dark zones be described as in equations 2.6 and 2.7. The subscript "b" and "d" indicates the bright and dark zones respectively.

$$E_\mathrm{b} = \|\mathbf{p}_\mathrm{b}\|^2 = \|\mathbf{H}_\mathrm{b}\mathbf{g}\|^2 \tag{2.6}$$

$$E_\mathrm{d} = \|\mathbf{p}_\mathrm{d}\|^2 = \|\mathbf{H}_\mathrm{d}\mathbf{g}\|^2 \tag{2.7}$$

Where $\mathbf{H}_\mathrm{b} = \mathbf{H}_1$, denotes the bright zone and $\mathbf{H}_\mathrm{d} = [\mathbf{H}_2^H, ..., \mathbf{H}_\mathrm{Q}^H]^H$ denotes the dark zone. In order to ensure that the desired sound energy is optimized simultaneously both for the bright and the dark zones, the optimization problem can be formulated so that the sound energy in the bright zone is maximised with the constraint of the energy in the dark zones being limited to a very small value D_0. An additional constraint to limit the power consumption of the loudspeaker array, the array effort, is added $\|\mathbf{g}\|^2 < E_0$. The array effort describes the maximum amount of energy that can be supplied to the loudspeakers. It can be an issue if too much energy is fed to the loudspeaker, because the loudspeaker is only linear in a given span, meaning that too much amplification could cause distortion which can be both linear or non-linear [10]. Thus, the array effort in such a system is preferred to be low, because the system is calculated as an linear-time-invariant (LTI) system. If this condition is not fulfilled, the system can not be assumed to behave as in the calculations. The constraint to the array effort also ensures that the leakage outside of the sound zones are not excessive and do in general make the implementation more robust both to driver positioning errors, and changes in the acoustic environment [1]. The acoustic contrast control optimization problem can then be written as:

$$\max_g \|\mathbf{H}_\mathrm{b}\mathbf{g}\|^2 \tag{2.8}$$

$$\text{subject to } \|\mathbf{H}_\mathrm{d}\mathbf{g}\|^2 \leq D_0 \tag{2.9}$$

$$\|\mathbf{g}\|^2 \leq E_0 \tag{2.10}$$

Using the Lagrangian this can be rewritten into an unconstrained optimization problem [11]:

$$\max_g L_c(\mathbf{g}) = \|\mathbf{H}_\mathrm{b}\mathbf{g}\|^2 - \lambda_1 \left(\|\mathbf{H}_\mathrm{d}\mathbf{g}\|^2 - D_0\right) - \lambda_2 \left(\|\mathbf{g}\|^2 - E_0\right), \qquad \lambda_1, \lambda_2 \geq 0 \tag{2.11}$$

Where λ_1 and λ_2 are the Lagrangian multipliers. These are used and scaled in order to determine the importance of each constraint. λ_1 controlling the constraint for the energy in the dark zone and λ_2 controlling the array effort. The solution that maximizes the Lagrangian is found by taking the derivative of L_c with respect to $\mathbf{g}$, and setting this equal to zero [8]:

$$\frac{\partial L_c(\mathbf{g})}{\partial \mathbf{g}} = 2(\mathbf{H}_\mathrm{b}^H\mathbf{H}_\mathrm{b}\mathbf{g} - \lambda_1\mathbf{H}_\mathrm{d}^H\mathbf{H}_\mathrm{d}\mathbf{g} - \lambda_2\mathbf{g}) = 0 \tag{2.12}$$

With this vector being equal to zero, it can be rewritten as a generalized eigenvector problem:

$$\mathbf{H}_{\mathrm{b}}^{H}\mathbf{H}_{\mathrm{b}}\mathbf{g} = \lambda_1 \left[\mathbf{H}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}} - \frac{\lambda_2}{\lambda_2} I\right]\mathbf{g} \tag{2.13}$$

Where $\mathbf{g}$ is chosen to be the eigenvector that corresponds to the largest eigenvalue of the matrix $[\mathbf{H}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}} + (\lambda_2/\lambda_1)I]^{-1}[\mathbf{H}_{\mathrm{b}}^{H}\mathbf{H}_{\mathrm{b}}]$ in order to maximize L_c. This can be rewritten to a form which later is comparable to the solution of pressure matching [12]:

$$[\mathbf{H}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}} + (\lambda_2/\lambda_1)I]^{-1}[\mathbf{H}_{\mathrm{b}}^{H}\mathbf{H}_{\mathrm{b}}]\mathbf{g} = \lambda_1\mathbf{g} \tag{2.14}$$

The Lagrangian multipliers are weightings to the constraints, where the ratio between them, $\lambda = \lambda_2/\lambda_1$, is the tradeoff between the performance and the array effort of the loudspeaker array. An important note is, the smaller the value gets, the less stable the system will become, since λ helps the matrix to not be ill conditioned, which can cause large numerical errors when doing a matrix inversion [8].

2.2.2 Pressure Matching

The aim of the pressure matching (PM) method is to reproduce the sound field in the bright zones at full strength while trying not to produce any signal in the other zones. The target sound field is described as $\mathbf{p}_{\mathrm{des}}$, which is the desired pressure that should be reproduced in the bright zone. The objective for the PM along with constraints on the sound energy in the dark zones and the array effort constraint is formulated by [1]:

$$\min_{g} \|\mathbf{H}_{\mathrm{b}}\mathbf{g} - \mathbf{p}_{\mathrm{des}}\|^2 \tag{2.15}$$

$$\text{subject to } \|\mathbf{H}_{\mathrm{d}}\mathbf{g}\|^2 \leq D_0 \tag{2.16}$$

$$\|\mathbf{g}\|^2 \leq E_0 \tag{2.17}$$

As in ACC, the problem can be rewritten into a single cost function using the Lagrangian. The cost function is given as:

$$\min_{g} L_p(\mathbf{g}) = \|\mathbf{H}_{\mathrm{b}}\mathbf{g} - \mathbf{p}_{\mathrm{des}}\|^2 + \lambda_1\left(\|\mathbf{H}_{\mathrm{d}}\mathbf{g}\|^2 - D_0\right) + \lambda_2\left(\|\mathbf{g}\|^2 - E_0\right), \quad \lambda_1, \lambda_2 \geq 0 \tag{2.18}$$

The solution which minimize the cost function in (2.18) is formulated in (2.20) and is obtained by taking the derivative of L_p with respect to $\mathbf{g}$, equal to zero:

$$\frac{\partial L_p(\mathbf{g})}{\partial \mathbf{g}} = 2\mathbf{H}_{\mathrm{b}}^{H}\left(\mathbf{H}_{\mathrm{b}}\mathbf{g} - \mathbf{p}_{\mathrm{des}}\right) + 2\lambda_1\mathbf{H}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}}\mathbf{g} + 2\lambda_2\mathbf{g} = 0 \tag{2.19}$$

$$\updownarrow$$

$$\left[\mathbf{H}_{\mathrm{b}}^{H}\mathbf{H}_{\mathrm{b}} + \lambda_1\mathbf{H}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}} + \lambda_2 I\right]\mathbf{g} = \mathbf{H}_{\mathrm{b}}^{H}\mathbf{p}_{\mathrm{des}} \tag{2.20}$$

To choose appropriate values for the Lagrangian multipliers, λ_1 and λ_2, equation (2.20) can be solved using an interior point algorithm [13]. A more simple approach can be made by setting

the parameter $\lambda_1 = 1$, which will give equal effort, to matching the pressure in the bright zone and minimizing the energy in the dark zone [1]. This solution can then be written as:

$$\left[\mathbf{H}_{\mathrm{b}}^{H}\mathbf{H}_{\mathrm{b}} + \mathbf{H}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}} + \lambda_2 I\right]^{-1}\mathbf{H}_{\mathrm{b}}^{H}\mathbf{p}_{\mathrm{des}} = \mathbf{g}_p \tag{2.21}$$

Where it can be seen that the only constraint is λ_2 which control the array effort and add some robustness, by avoiding ill-conditioned problems [1]. It can also be seen that, with $\lambda_1 = 1$, equation (2.21) is identical to the ACC (equation (2.14)) if the following two criterias are met:

1. The target pressure $\mathbf{p}_{\mathrm{des}}$ in the bright zone should be an ACC solution, given that $\mathbf{p}_{\mathrm{des}} = \mathbf{H}_{\mathrm{b}}\mathbf{g}_{\mathrm{c}}$, where the subscript "c" is used for indicate ACC.
2. The constrains for D_0 and E_0 should be identical.

A combination of the two techniques has also been proposed in [3, 14]. The optimal solution is given by:

$$\left[\xi\mathbf{H}_{\mathrm{b}}^{H}\mathbf{H}_{\mathrm{b}} + (1-\xi)\mathbf{H}_{\mathrm{d}}^{H}\mathbf{H}_{\mathrm{d}}\right]^{-1}\xi\mathbf{H}_{\mathrm{b}}^{H}\mathbf{p}_{\mathrm{des}} = \mathbf{g}_{\mathrm{cb}} \tag{2.22}$$

Where the value ξ is a weighting factor that determines the balance between the energy in the dark zone and the mean square error (MSE) in the bright zone [14]. The value ξ should be within $(0 \leq \xi < 1)$ [3, 14]. When ξ goes to 0, the solution approximates the ACC and when ξ goes to 1, the solution approximates the PM [14].

2.2.3 Comparison Between Acoustic Contrast Control and Pressure Matching

ACC always gives a high contrast level between the zones with a low array effort. However, the downside is that the sound field reproduction would not be as accurate for the listener [1]. This is due to the optimization problem that tries to maximize the difference between the zones with no constraints to the target sound field. However in PM, part of the optimization problem is to retain the desired signal, and thus provide a better reproduction of the signal compared to the ACC where the reproduction error is higher, due to the goal solely being to maximize the acoustic contrast [1].

Both of these algorithms assume knowledge of the acoustic transfer functions from the loudspeaker to each of the control points. In an anechoic environment this is quite simple and can easily be estimated as shown in section 2.1. However, in a real application this is not always the case. The transfer function will here include reflections, which are dependent on the the room shape, including the walls, windows, the objects within it and the temperature. These parameters will be discussed later in section 2.4.

In the next section the state of the art algorithms will be explained, where one will be chosen and tested with different settings.

2.3 State of the Art Beamforming Algorithms

When using beamforming for personal sound zones, different algorithms are proposed, where they in general are based on either acoustic contrast control (ACC) or pressure matching (PM). The different algorithm have different strengths in order to solve some of the limitations of beamforming, often by adding more constrains to the optimization problem. E.g. A. Canclini et al. have in [15] tried to take the room into account by adding a simple room model as a part of the optimization problem, where different orders of reflection can be used. Also, the difference by solving the optimization problem in the time domain or the frequency domain should be determine, where e.g [4, 15] are solved in the frequency domain and Marcos F. Simón Gálvez et al. gives a method in [3] to do the calculations in the time domain. The study also include the amount of samples needed to calculate the filters. A downside by doing the optimization in the time domain is, that the dimensions of the problem can become large relatively quickly compared to do the independent optimization in the frequency domain [5]. However, to create the filter, as an impulse response, filter weights calculated in the frequency domain can be truncated when using an inverse fast Fourier transform (IFFT) [3, 5], due to the active window size of the transform.

In the following subsection an algorithm proposed by Vicent Molés-Cases et al. in [4] will be described. The algorithm is called *Weighted Pressure Matching Total Variation* (wPM-TV), which is based on a hybrid between the acoustic contrast control and pressure matching as described in equation (2.22). This algorithm is chosen for further examination because it allows the possibility to determine the balance between the two. Also, the TV part of the algorithm aims to make a more uniform sound image in the dark zone, which will be described in details later.

2.3.1 Weighted Pressure Matching With Total Variation

The Weighted Pressure Matching with Total Variation (wPM-TV) is an algorithm proposed in [4], which deal with one of the limitations of the pressure matching algorithm. When trying to minimize the mean acoustic potential energy in the dark zone one problem is that the sound field is not necessarily uniform, some of the control points in the dark zone can have high acoustic energy with others having low acoustic energy, while still having an overall low mean acoustic energy [4]. If the acoustic energy levels are very uneven within the dark zone, this can be disturbing for a listener located within the dark zone, due to the sudden changes in energy [4]. Graph Signal Processing (GSP) is used to apply an additional constraint in order to take the spatial uniformity of the acoustic potential energy in the dark zone into account. A block diagram of the algorithm and which parameters that can be tweaked can be seen in figure 2.4. This diagram corresponds to the ”algorithm block” in figure 2.3. The different parameter will be described in the following.

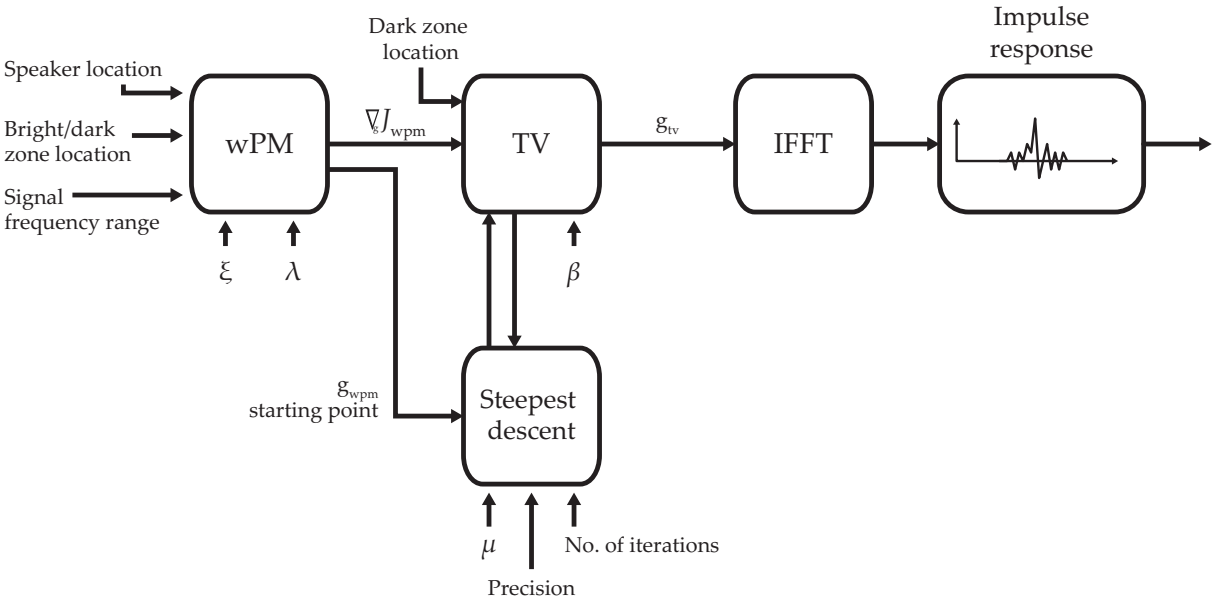


Figure 2.4: Block diagram showing the signal in the wPM-TV algorithm.

wPM

The system consist of an array of L loudspeakers and two regions: the bright zone with M_b control points, and the dark zone with M_d control points. The acoustic pressure in a the m'th control point in zone q can be described as in (2.23). $q = \{b, d\}$ and describe the total amount of zones, where the subscript b is used for the bright point and d is used for the dark points.

$$p_{qm} = \sum_{l=1}^{L} h_{qm,l} \cdot g_l = \mathbf{h}_{qm}\mathbf{g} \tag{2.23}$$

where $\mathbf{g}$ is a vector which describes the source strength for all L loudspeakers to a given frequency. The vector $\mathbf{h}_{qm}$ containing the acoustic transfer functions (ATF) for the L loudspeakers and the m'th control point. For all q zones, the ATF's can be described as:

$$\mathbf{H}_q = \begin{bmatrix} \mathbf{h}_{q1} \\ \vdots \\ \mathbf{h}_{qM_q} \end{bmatrix} = \begin{bmatrix} h_{q1,1} & \dots & h_{q1,L} \\ \vdots & \ddots & \vdots \\ h_{qM_q,1} & \dots & h_{qM_q,L} \end{bmatrix} \tag{2.24}$$

The ATF between the L loudspeakers and the m'th measuring point can be expressed as the free-field propagation model which was described in section 2.1:

$$h_{qm,l}(\omega) = \frac{e^{-j\omega \frac{\|\Delta_{qm,l}\|}{c}}}{4\pi \|\Delta_{qm,l}\|} \tag{2.25}$$

Where $\Delta_{qm,l}$ is the distance between a loudspeaker l and a measuring point qm, $\omega = 2\pi f$ and c is the speed of sound in air.

For each frequencies, ω, of interest, the aim of the algorithm is to find the source strength vector $\mathbf{g}$ that minimize the cost function given in (2.26):

$$J_{\text{wpm}}(\mathbf{g}) = \xi\epsilon_b(\mathbf{g}) + (1-\xi)E_d(\mathbf{g}) + \lambda \left\|\mathbf{g}\right\|^2 \tag{2.26}$$

Where $0 < \xi < 1$ is a weighting factor as described in (2.22) and λ is a regularization parameter which is used to control/limit the energy of $\mathbf{g}$ - the array effort. E_d and ϵ_b are the mean acoustic potential energy for the dark zone and the mean square error (MSE) for the bright zone respectively. These are given by (2.27) and (2.28)

$$E_d = \frac{1}{M_d}\left\|\mathbf{H}_d\mathbf{g}\right\|^2 = \frac{1}{M_d}\mathbf{g}^H\mathbf{H}_d^H\mathbf{H}_d\mathbf{g} \tag{2.27}$$

$$\epsilon_b = \frac{1}{M_b}\left\|\mathbf{H}_b\mathbf{g} - \mathbf{d}_b\right\|^2 = \frac{1}{M_b}\left(\mathbf{H}_b\mathbf{g} - \mathbf{d}_b\right)^H\left(\mathbf{H}_b\mathbf{g} - \mathbf{d}_b\right) \tag{2.28}$$

Where $\mathbf{d}_b$ is a vector of the target sound field in the m'th control point in the bright zone. As described in the former section, the optimal source strength vector $\mathbf{g}$ is found by taking the gradient of the cost function. The gradient of the cost function (2.26) for wPM is given by [4]:

$$\nabla_{\mathbf{g}} J_{\text{wpm}}(\mathbf{g}) = 2\left[\frac{\xi}{M_b}\left(\mathbf{H}_b^H\mathbf{H}_b\mathbf{g} - \mathbf{H}_b^H\mathbf{d}_b\right) + \frac{(1-\xi)}{M_d}\mathbf{H}_d^H\mathbf{H}_d\mathbf{g} + \lambda\mathbf{g}\right] \tag{2.29}$$

Where the optimal solution is when $\nabla_{\mathbf{g}} J_{\text{wpm}}(\mathbf{g}) = 0$, and a closed solution is given by [4]:

$$\mathbf{g}_{\text{wpm}} = \left(\frac{\xi}{M_b}\mathbf{H}_b^H\mathbf{H}_b + \frac{(1-\xi)}{M_d}\mathbf{H}_d^H\mathbf{H}_d + \lambda\mathbf{I}\right)^{-1}\frac{\xi}{M_b}\mathbf{H}_b^H\mathbf{d}_b \tag{2.30}$$

To see the performance of the algorithm, simulations with a simple setup have been made. The simulations also include the influence of ξ, to see the difference between ACC, PM and a mix of the two. When $\xi = 0$ the optimal solution will be ACC and when $\xi = 1$ it will be PM. However, if $\xi = 0$ the optimal solution will be $\mathbf{g} = 0$, meaning that the optimal solution will be to just turn off the loudspeakers, which seen from the algorithms point of view will be infinite contrast. Therefore, ξ is set to a small value in the simulation, 1e-9. The results can be seen in figure 2.5, and the settings are: $\omega = 1$ kHz, 29 loudspeakers spaced 0.05 m from each other and are simulated as point sources, λ =1e-5, one bright zone placed in $[0.75, 1]$ and one dark zones placed in $[-0.75, 1]$. It should be noted that λ =1e-5 will increase the array effort, thus in a real application, the loudspeakers might distort if the array effort is too high, and λ needs to be chosen to avoid that from happening. The value of λ in this case is chosen for illustration purpose.

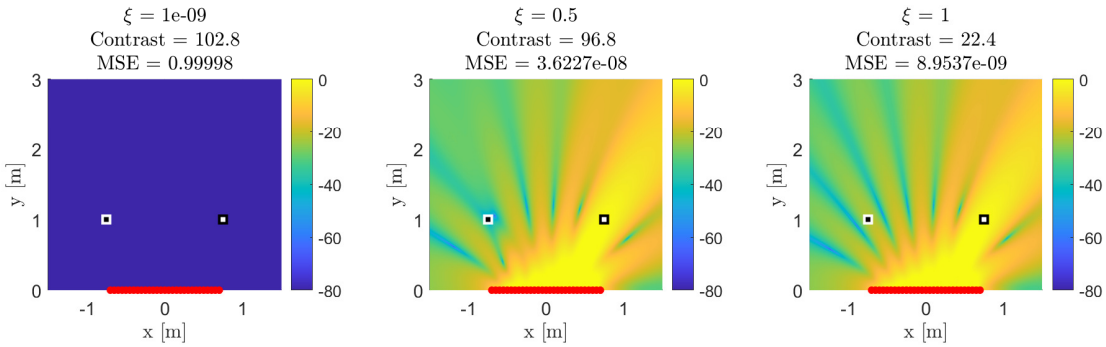


Figure 2.5: Simulation of wPM algorithm, which show the impact of the tuning parameter ξ, with a fixed regularisation parameter λ =1e-5. 29 loudspeakers are used spaced 0.05 m from each other with center in $[0, 0]$, a bright zone placed in $[0.75, 1]$ and a dark zone placed in $[-0.75, 1]$. The plot is with the frequency set to 1 kHz. The contrast is the difference between the bright zone and the dark zone. The color bar is given in normalized dB.

As it can be seen in figure 2.5 ACC provide higher contrast than PM and the mixed is in between. However when $\xi = 1e - 9$ the amplitude is so low that there is effectively no signal, this is also indicated in the high reproduction error. so even though the contrast is higher, this would provide very bad experience for the listener. Changing the array effort regularization parameter λ to another value in the same scenario, the difference in contrast is not as big. This can be seen in figure 2.6 with $\lambda = 0.1$. However, the contrast between the two zones decreases as the value of λ increases, Thus a lambda value that assures stability should be found before implementing it in a real system. Also it can be seen that the MSE increases, which make sense since there are less energy to reproduce the signal.

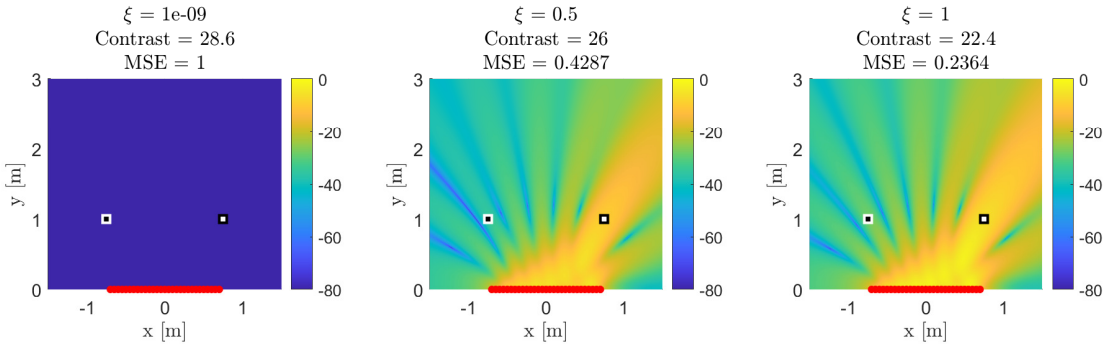


Figure 2.6: Simulation of wPM algorithm, which show the impact of the tuning parameter ξ, with a fixed regularisation parameter $\lambda = 0.1$. 29 loudspeakers are used spaced 0.05 m from each other with center in $[0, 0]$, a bright zone placed in $[0.75, 1]$ and a dark zone placed in $[-0.75, 1]$. The plot is with the frequency set to 1 kHz. The contrast is the difference between the bright zone and the dark zone. The color bar is given in normalized dB.

To compare the results of the wPM algorithm the same setup is simulated using the DS-beamformer. The results can be seen in figure 2.7.

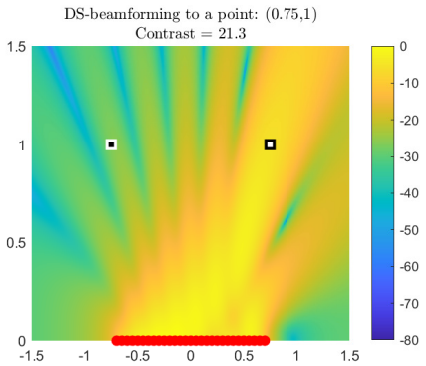


Figure 2.7: Simulation of DS-beamformer. 29 loudspeakers are used spaced 0.05 m from each other with center in $[0, 0]$, a bright zone placed in $[0.75, 1]$ and a dark zone placed in $[-0.75, 1]$. The plot is with the frequency set to 1 kHz. The contrast is the difference between the bright zone and the dark zone. The color bar is given in normalized dB.

When comparing the wPM algorithm it can be seen that the performance of the two algorithms is very similar when $\xi = 1$, with difference between the contrast of the wPM and the DS-beamformer being only 1.1 dB. This is because, as mentioned earlier, wPM with $\xi = 1$ becomes PM, meaning that the optimization is mainly reducing the MSE in the bright zone, and only uses a low effort towards minimizing the pressure in the dark zone. The DS-beamformer is not solving an optimization problem but is instead based on delays calculated to ensure that the phase align for each loudspeaker in the desired point, meaning that no effort is being used to reduce the pressure in the dark zone, and therefore provide very similar results as PM. However a slight change in ξ will improve the contrast even if the zones are placed less optimally.

In order to show how a small change in ξ affects both the contrast and the MSE a test is made, where ξ is increased by 0.1 starting from 1e-9 to 1. The results can be seen in figure 2.8. The plot is shown for $f = [300, 1000, 1500, 2000, 3000]$ Hz. The changes in the MSE's are, in this case, so small that they are all placed on top of each other.

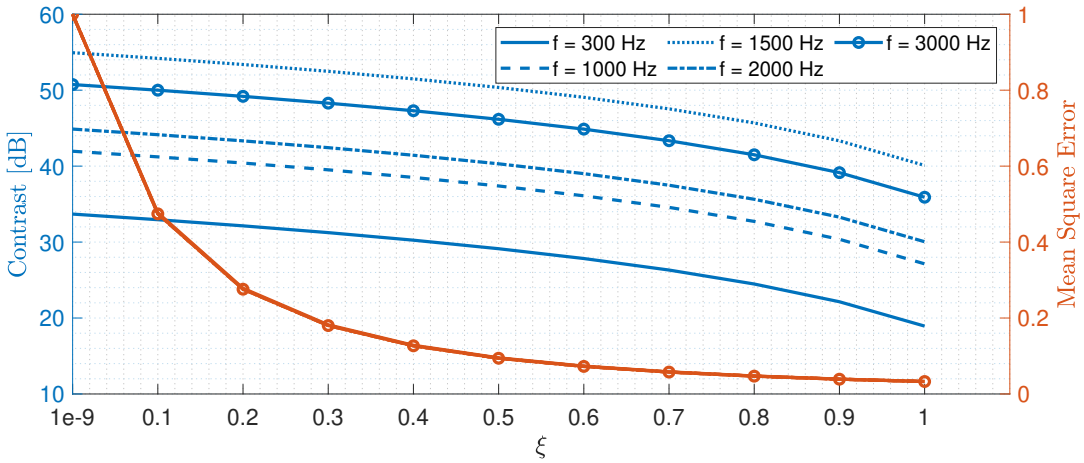


Figure 2.8: The contrast and MSE as function as different values of ξ for $f = [300, 1000, 1500, 2000, 3000]$ Hz, with $\lambda = 0.1$.

As it can be seen in figure 2.8 a quick decrease in the MSE is seen at low ξ values, where the change is much smaller at the higher values of ξ. The contrast seems to decrease at a more constant rate up until the higher ξ values, where this is decreased even further.

Figure 2.9 show the relationship between the array effort regularization parameter λ and the contrast between the zones, as well as its impact on the MSE. The plot is again shown for for $f = [300, 1000, 1500, 2000, 3000]$ Hz and $\xi = 0.9$. It should once again be noted that decreasing the value of λ increases the array effort which can cause distortion in a final system. This will be investigated and described in more details later in the report. The changes in the MSE's are again, in this case, so small that they are all placed on top of each other.

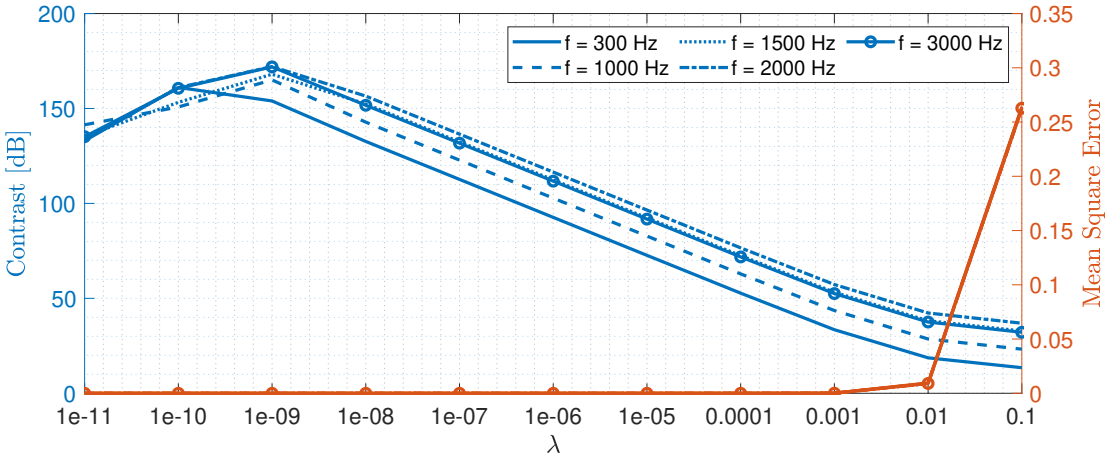


Figure 2.9: The contrast and MSE as function as different values of λ for $f = [300, 1000, 1500, 2000, 3000]$ Hz, with $\xi = 0.9$.

Finally the array effort for this setup will be examined. This will be done in order to see how the array effort behaves with different values of λ and different array sizes. The array effort can be calculated by equation (2.31), which is the total energy fed to the array relative to the pressure in the bright zone, p_r [16].

$$\text{Effort} = 10 \log_{10} \left(\frac{\mathbf{g}_{\text{wpm}}^H \mathbf{g}_{\text{wpm}}}{|p_r|^2} \right) \tag{2.31}$$

The results can be seen in figure 2.11, where the decreasing of the array are towards the middle. The arrays can be seen in figure 2.10.

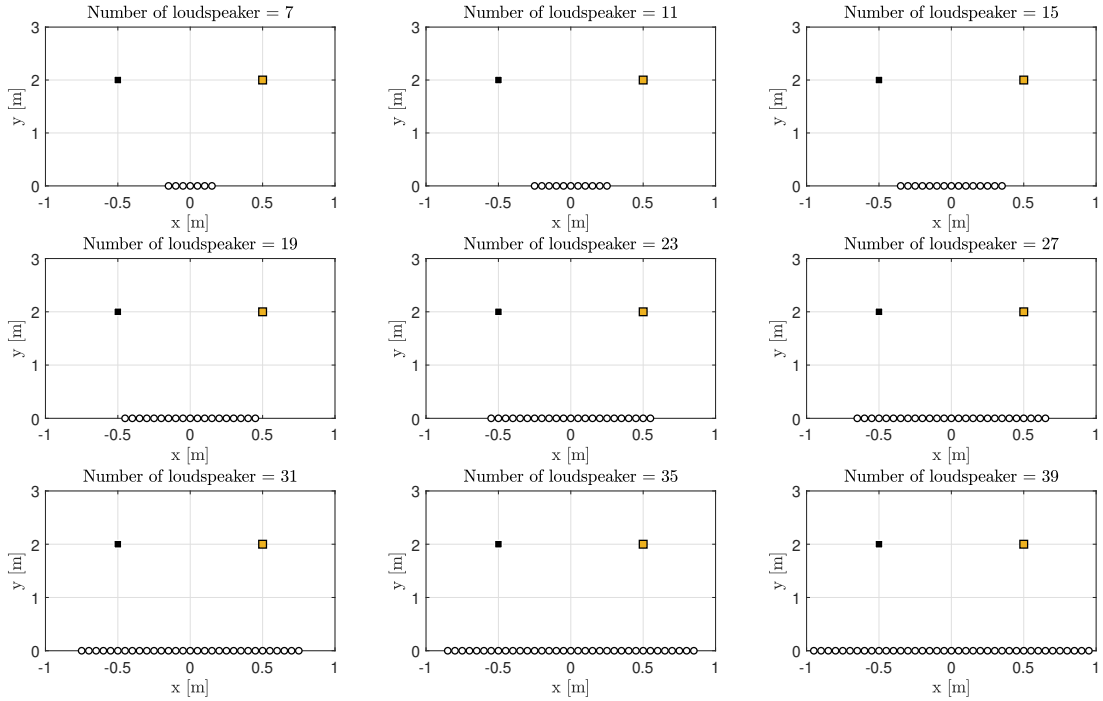


Figure 2.10: Illustration of the different array sizes and the placement of the bright (yellow) and dark (black) zones. The loudspeakers are space 0.05 m from each and the center of the array is in $[0, 0]$.

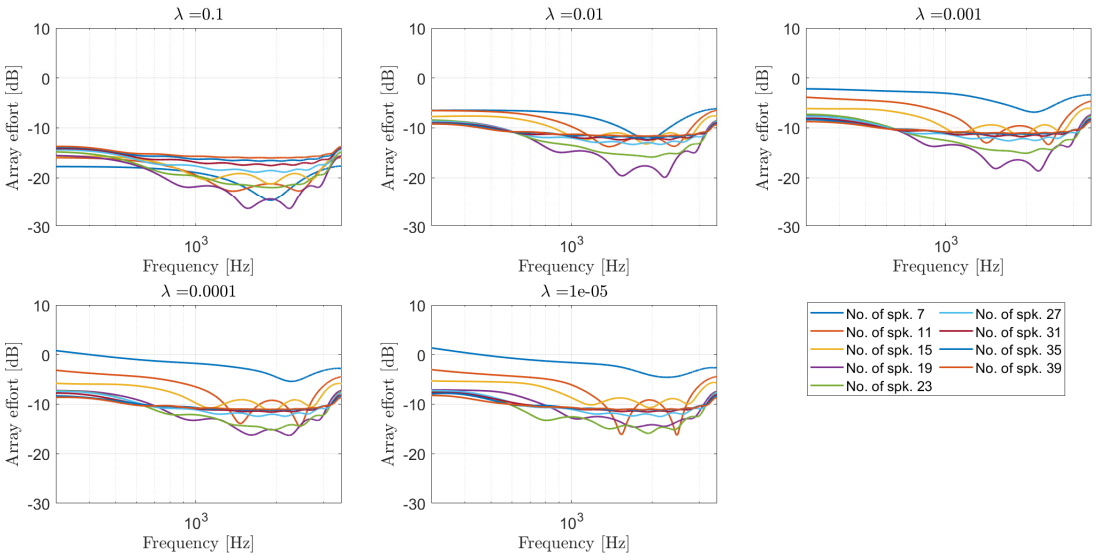


Figure 2.11: Array effort as function of frequency with different λ's and amount of loudspeakers.

As it can be seen in figure 2.11, the array effort depends on the amount of loudspeakers in the array which becomes more and more clear as the value of λ decreases. When having more loudspeakers, the array effort is close to the same when the value of λ changes, since more loudspeakers can be used to reproduce the desired signal, and therefore the energy needed can be distributed to more loudspeakers. When the array effort becomes higher, it implies poor

acoustical efficiency, where high sound pressure levels are emitted in to the room [16].

On a last note, these results will vary from setup to setup and frequency to frequency, which will be clarified in the next chapter.

wPM-TV

Using equation (2.30), the solution can minimize the acoustic potential energy in the dark zones. This will not necessarily be true if there are used more control points. The algorithm is therefore expanded, by adding a new constrain, to improve the spatial uniformity. As earlier mentioned GSP is used in order to include the spacial uniformity of the acoustic potential energy of the dark zone in the optimization problem.

A graph is used to characterize the relation between the acoustic potential energy in the different control points in the dark zone. A graph refers to graph theory in which it is used to describe a mathematical structure. A graph is defined as $\mathcal{G} = (\mathcal{V}, \mathcal{A})$ [17], where $\mathcal{V}$ is a set of vertices of $\mathcal{G}$, which in this context is associated with the control points in the dark zone, resulting in a total of M_d vertices defined as $\mathcal{V} = 1, ..., M_d$. $\mathcal{A} \in \mathbb{R}^{M_d \times M_d}$ is a subset of $\mathcal{P}_2(\mathcal{V})$, meaning that the edge, $\mathcal{A}$, is a set of two-element subsets of $\mathcal{V}$ describing the similarity between the vertices [4]. This is illustrated in figure 2.12.

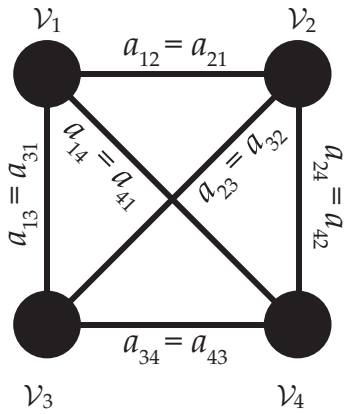


Figure 2.12: Illustration of the similarity of the vertices with four control points in the dark zone, $M_d = 4$.

In the wPM-TV algorithm, the proposed similarity is based on the distance between the control points within the dark zone, with reasoning being that it reflects the physical relation between the control points [4]. The elements of the $\mathcal{A}$ is defined as:

$$a_{ij} = e^{\frac{-\left\|\mathbf{z}_i - \mathbf{z}_j\right\|^2}{2\sigma^2}} \tag{2.32}$$

Where: $\mathbf{z}_i \in \mathbb{R}^3$ is the spatial coordinates for each control point and σ is given by:

$$\sigma = \frac{2}{M_d(M_d - 1)} \sum_{i=1}^{M_d} \sum_{j=i+1}^{M_d} \|z_i - z_j\| \tag{2.33}$$

Which is the mean distance between all of the control points within the dark zone. The graph signal is defined as a column vector with the signal for each of the vertices. In this case, the pressure for each of the control points are defined as: $f_{dm} = p_{dm} = \sqrt{\mathbf{g}^H \mathbf{h}_{dm}^H \mathbf{h}_{dm} g}$. The graph signal can then be used to calculate the Total Variation (TV) of the graph which can be used as an indicator of the uniformity of the sound field in the dark zone [4]. The TV of the graph is defined as [18]:

$$T(\mathbf{f}_d) = \mathbf{f}_d^T \mathbf{L} \mathbf{f}_d = \sum_{i=1}^{M_d} \sum_{j=1}^{M_d} l_{ij} f_{di} f_{dj} = \frac{1}{2} a_{ij} (f_{di} - f_{dj})^2 \tag{2.34}$$

This TV is then included in the cost function, as a constraint to optimize towards a uniform acoustic field in the dark zone. The new cost function, which is used in the wPM-TV algorithm is then defined as:

$$J_{\text{tv}}(\mathbf{g}) = J_{\text{wpm}}(\mathbf{g}) + \beta T(\mathbf{f}_d(\mathbf{g})) = \xi \varepsilon_b(\mathbf{g}) + (1 - \xi) E_d(\mathbf{g}) + \lambda \left\| \mathbf{g} \right\|^2 + \beta T(\mathbf{f}_d(\mathbf{g})) \tag{2.35}$$

Where: $\beta \in \mathbb{R}^+$ is the weighting of the TV of the graph in the cost function. The higher value of β the higher uniformity in the dark zone. In order to find an optimal solution, g, for the new cost function, the gradient of the cost function $J_{\text{wpm}}(\mathbf{g})$ with respect to $\mathbf{g}$ is found.

$$\nabla_{\mathbf{g}} f_{dm}(\mathbf{g}) = \nabla_{\mathbf{g}} \left(\sqrt{\mathbf{g}^H \mathbf{h}_{dm}^H \mathbf{h}_{dm} \mathbf{g}} \right) = \frac{1}{f_{dm}(\mathbf{g})} \mathbf{h}_{dm}^H \mathbf{h}_{dm} \mathbf{g} \tag{2.36}$$

And from this the Jacobian matrix of $f_d(\mathbf{g})$ with respect to $\mathbf{g}$ can be written as:

$$\mathbf{J}(\mathbf{f}_d(\mathbf{g})) = [\nabla_{\mathbf{g}} f_{d1}, \ldots, \nabla_{\mathbf{g}} f_{dM_d}(\mathbf{g})]^T \tag{2.37}$$

By using equation (2.34) and (2.37) the gradient of the Total Variance can be expressed as in equation (2.38) [4]:

$$\nabla_{\mathbf{g}} T(\mathbf{f_d}(\mathbf{g})) = \nabla_{\mathbf{g}} \left(\sum_{i=1}^{M_d} \sum_{j=1}^{M_d} l_{ij} f_{di} f_{dj} \right) = 2 \mathbf{J}(\mathbf{f}_d(\mathbf{g}))^T \mathbf{L} \mathbf{f}_d(\mathbf{g}) \tag{2.38}$$

Combined with the gradient of $J_{\text{wpm}}(\mathbf{g})$ found in equation (2.29), the gradient of the new cost function with the uniformity of the dark zone included (2.35) can be derived as:

$$\nabla_{\mathbf{g}} J_{\text{tv}}(\mathbf{g}) = \nabla_{\mathbf{g}} J_{\text{wpm}}(\mathbf{g}) + 2\beta 2 \mathbf{J}(\mathbf{f}_d(\mathbf{g}))^T \mathbf{L} \mathbf{f}_d(\mathbf{g}) \tag{2.39}$$

However since a analytical solution $\mathbf{g}$ for $\nabla_{\mathbf{g}} J_{\text{tv}}(\mathbf{g}) = 0$ can not be found according to [4], steepest descent is then used to find the optimal values for the vector $\mathbf{g}$. The steepest descent algorithm is given by [19] as:

$$\mathbf{g}_{\text{tv}}^{k+1} = \mathbf{g}_{\text{tv}}^k - \mu \nabla_g J_{\text{tv}}(\mathbf{g}_{\text{tv}}^k) \tag{2.40}$$

Where μ is the step size of the steepest descent algorithm and k is the iteration number. By choosing the initial $\mathbf{g}_{\text{tv}}^k = \mathbf{g}_{\text{wpm}}$ to be the optimal solution of wPM the local minimum found is

the closest to the solution of the global minimum of the wPM cost function, meaning that the solution does not change considerable compared to the wPM solution, but is still improving the uniformity in the dark zone [4].

To test the TV-part of the algorithm, a simulation is made. The loudspeakers placement are the same as in figure 2.5, but this time, a grid of control points are used in the dark zone. This is used in order to make the algorithm make the graphs between the points, and make the dark zone more uniform for all the control points. The grid is 3×3 with the points spaced 0.05 m with center in $[-0.45, 1]$, so that it covers an area of 10×10 cm. The bright zone is placed in $[0.5, 1]$, and the testing frequency is at 1 kHz. The results can be seen in figure 2.14. The regularization parameter $\lambda = 0.1$, and $\xi = 0.9$.

When performing the wPM-TV, the steepest descent algorithm is used, since it is a non-convex problem. The steepest descent algorithm can be written i MATLAB as in code example 2.1, where the values showed is also those used in the simulation showed in figure 2.14.

Code example 2.1: Steepest descent implemented in MATLAB.

```
stepsize = 0.1;
max_iterations = 10000;
g_new = gwpm;
precision = 2^(-23);
for i = 1:max_iterations
    g_old = g_new;
    g_new = g_old - stepsize * Grad_Jtv(g_old);
    if max(abs(g_new - g_old)) < precision
        break;
    end
end
```

The `max_iterations` ensure that the algorithm does not run forever if the precision is never reached. The step size is chosen because it have shown fine results and have a good relation between speed and precision. Finally, the precision is chosen because of the bit-resolution that high quality music is played with, which is 24-bit. In order to see the mean square error (MSE) as function of iterations, test with different values of β have been done. Additionally the difference in the filter coefficients are plotted to see the behaviour of the steepest descent seen in figure 2.13. Here it can be seen that the changes in the coefficients becomes smaller with the number of iterations. The reason why the plots have different lengths, is because of the stopping/precision criteria.

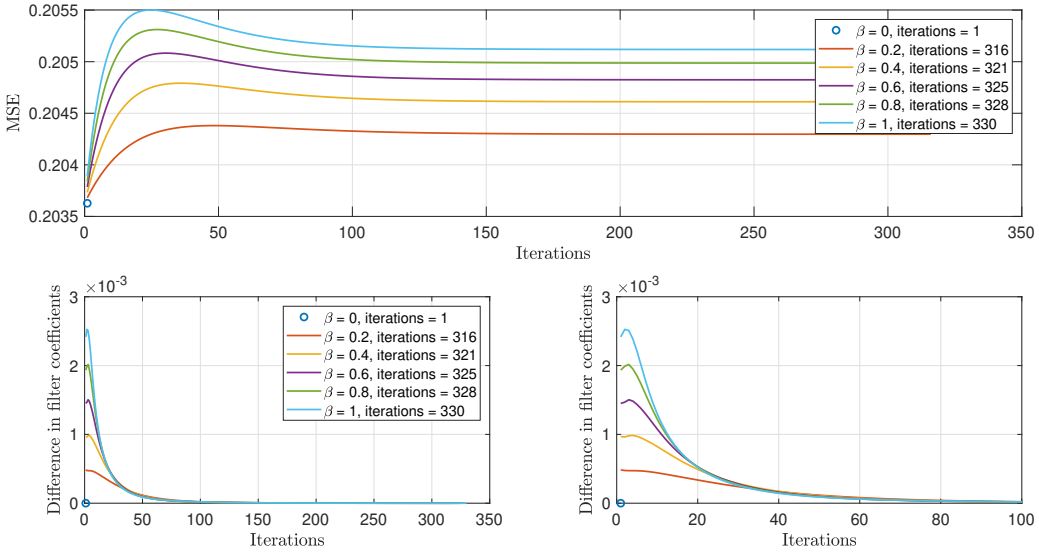
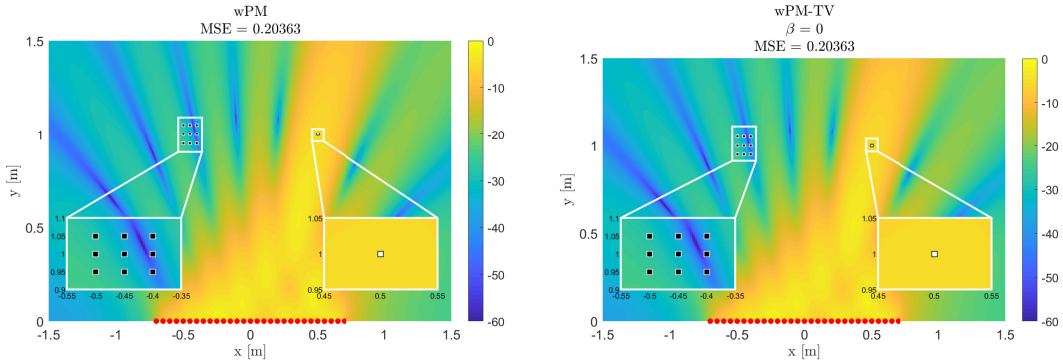
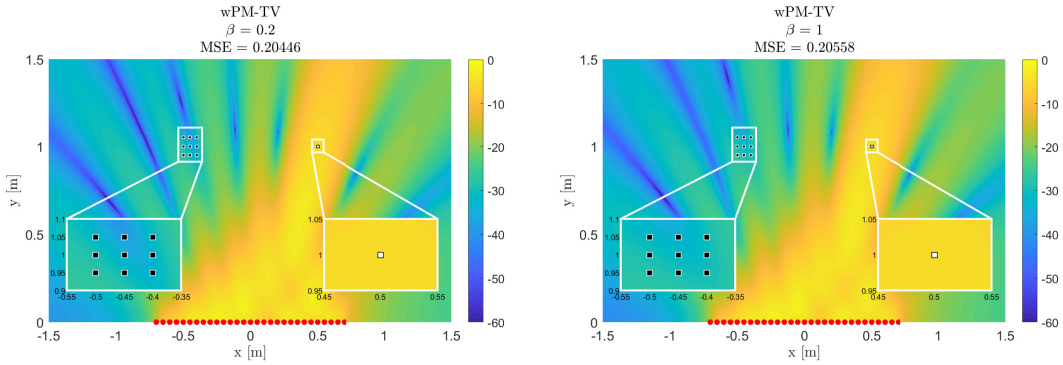


Figure 2.13: Simulation of the behavior of the MSE and difference in the filter coefficients as function of iterations in the steepest descent algorithm with different values of β.

As it can be seen in figure 2.13 the MSE increases with a higher value of β. Also it can be seen that the difference in the filters from iteration to iteration behaves as expected. It can be seen that the higher value of β, the more iterations are needed before the stopping criterion is met. It may be due to the specific problem and the step size.



(a) wPM **(b)** wPM-TV: $\beta = 0$



(c) wPM-TV: $\beta = 0.2$ **(d)** wPM-TV: $\beta = 1$

Figure 2.14: Simulation of wPM-TV algorithm, which show the impact of parameter β. $\xi = 0.995$ and the regularization parameter $\lambda = 0.1$. 29 loudspeakers are used spaced 0.05 m from each other with center in 0, a bright zones placed in $[0.5, 1]$ and a 3×3 grid of dark zones with center in $[-0.45, 1]$, separated 0.05 m from each other. The plot is with the frequency set to 1 kHz.

It can be seen in figure 2.14, that the higher value of β the more uniform the dark area becomes. Also it can be seen that when $\beta = 0$, it is similar to wPM as expected. By choosing a smaller step size, μ, when β gets higher the amount of iterations requires can be reduced. In this simulation, the figure 2.14b stopped after 1 iteration, figure 2.14c stopped after 308 iterations, and figure 2.14d reached the precision after 329.

However choosing the right step size is also important, both to ensure fast convergence but also to avoid making the system unstable and unable to find a sufficient solution. A small step size ensures that a sufficient solution can be found, where a bigger step size often result in a faster convergence. However if the chosen step size is too large, the algorithm is unable to find a sufficient solution. In figure 2.15, the amount of iterations needed for different β's and step sizes, μ, can be seen, for two different setups of the bright and dark zones. The maximum amount of iterations was 10,000 meaning that if this amount was reached the algorithm was stopped and did not find a sufficient precise solution.

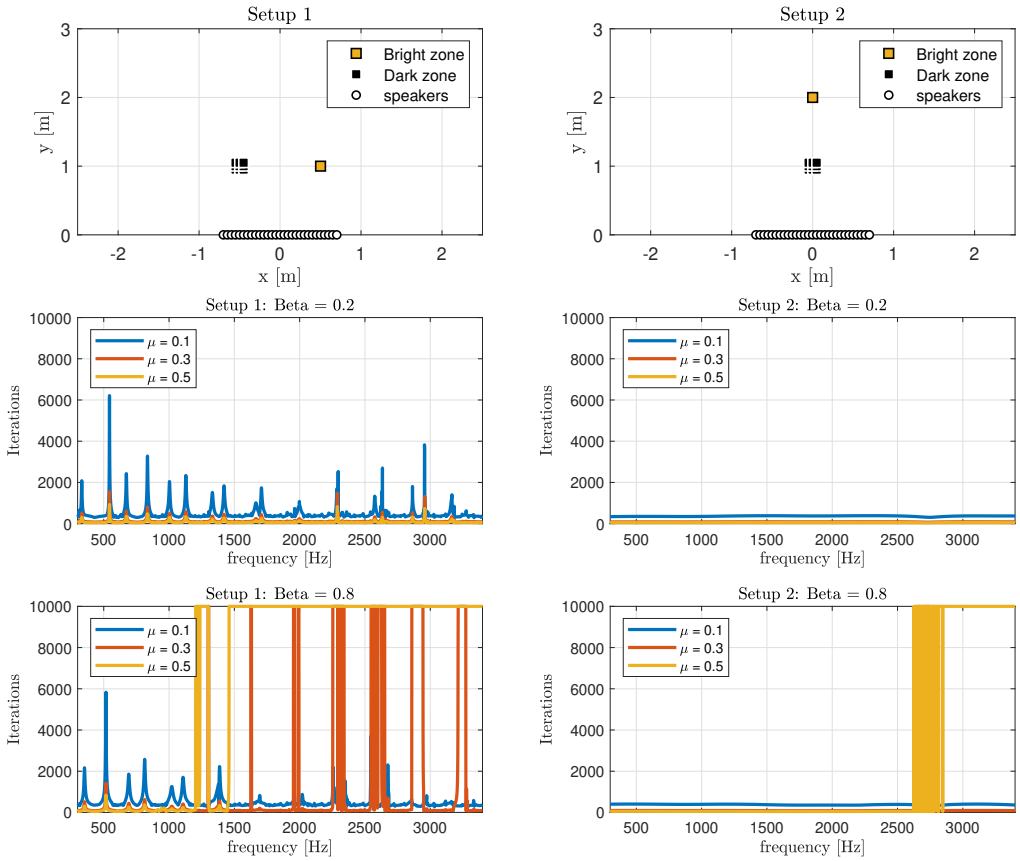


Figure 2.15: Iterations needed to find a solution, showed for different step sizes μ and weighting parameters β. $\xi = 0.995$ and the regularization parameter $\lambda = 0.1$. 29 loudspeakers are used spaced 0.05 m from each other with center in $[0, 0]$. The first setup with a bright zone placed in $[0.5, 1]$ and a 3×3 grid of dark zones, placed with center in $[-0.5, 1]$, separated 0.05 m from each other. The second setup with a bright zone placed in $[0, 2]$ and a 3×3 grid of dark zones, placed with center in $[0, 1]$, separated 0.05 m from each other.

Here it can be seen that the amount of iteration needed is highly dependent on the position of the dark and bright zones as well as the β value. Here it can be seen that when $\beta = 0.2$, the first setup uses significantly more iterations than the second setup. When $\beta = 0.8$ it can be seen that for both setups, the algorithm are having trouble finding solutions with $\mu = 0.5$.

2.4 Robustness of the Algorithms Used for Personal Sound Zones

When implementing the algorithm described in the section 2.3, different considerations should be taken into account, so that the implementation is not very sensitive to small changes in the acoustic transfer function, requiring new filters to be calculated. When beaming in a direction, far-field assumptions are used for the optimization. This does however not mean that this would be the same for a real room. E.g. reflections can have an impact on the beamforming

performance. Meaning that if a side lobe with high energy hit a wall, reflections will be radiated into the room, where the reflected energy is determined by the reflection coefficient of the material. Measurements done by Simón Gélvez M. et al. described in in [20] showed that the extra pressure component, due to reflections, lowered the performance of the beamforming. Also by increasing the reverberation time of the room, the contrast between the two zones were lowered. However, the robustness of a line array was also tested. Here it was shown that small mismatches in the transfer functions leads to high reduction in the free field directivity, but does not have as high reduction in the performance when the directivity was measured [20].

The deviation in the placement of the loudspeakers and changes in temperature and their impact on the performance have been investigated by Coleman P. et al. described in [21]. The deviations of the loudspeakers placement were tested with 10 mm random deviation and showed that both ACC and PM is sensitive to the changes in the placement. The temperatures tested was with a change in temperature of 17°C. This corresponds to a change in the speed of sound of 10 m/s. According to [21] the results showed that only PM is sensitive to the temperature changes.

Another parameter that can have impact on the performance, is how sensitive the system is to small variations in magnitude and phase in the loudspeakers responses [5]. Therefore, the system can preferably be simulated with these errors to test the robustness of the system.

2.5 Evaluation and Comparison of Results

In order to compare the results a common performance indicator is required. The simplest one to use is simply achieved by taking the mean of the control points in the dark zone and see how it compares with the bight zone. However this does not take the listening experience into account, and this does not take the uniformity of the dark zone into account, which is why this is not in every case the best comparison.

A big challenge when it comes to the evaluation of the performance of sound zone control, is that the performance is both evaluated based on how well a system is able to create a contrast between the personal sound zones, but is also measured on the ability to be able to create the desired signal in the bright zone. These measurements are, however, often achieved with a trade-off between minimizing the mean square error and maximizing the contrast. This is seen in the wPM algorithm (2.26), where the ϵ parameter, determines whether an accurate reproduction of the sound field, a high contrast between the zones or a mix of the two, are preferred.

So in order to make an actual performance evaluation, some kind of measure of how an increase in MSE compares to a higher contrast, in terms of listening experience both for the listener in the bright zone but also the listener seated in the dark zone is needed.

2.5.1 Perceptual Evaluation of Contrast Between Sound Zones

When evaluating the results, an important part is how people perceive sound. The importance of this aspect is to give an idea of how high contrast is needed when creating sound zones, and also how the sound quality is. The target sound quality have been investigated by Baykaner K. et al. described in [22], by listing tests, between different sound zone methods, including

ACC and PM, as described earlier. The results showed that ACC provided a higher contrast compared to PM, but PM had a more accurate representation of the sound field. The results also showed that the sound quality was dependent on the quality of the reproduced sound and the distraction from interfering sounds [22].

In [23], the target to interferer ratio (TIR) has been investigated. It was a comparison between speech and music, and the results showed, that when TIR was above 25 dB the variance between the test subjects was low and the participants seemed to have reached a consensus that the scenarios were not distracting anymore [23]. However below the 25 dB, the results had a larger variance, which is probably a results of when people think something is distracting or not. However, it should be noted that this values is only a guideline. When the two signals shares a lot of the same frequency content, more masking [24] will occur and lower TIR-values will be acceptable.

However this does not tell anything about the perceptual difference between the original signal and reproduced signal. For this, an informal listening test done by the authors will be made.

2.6 Summary

Personal sound zones can be achieved by the use of beamforming described as an optimization problem. General optimization problems which forms the foundation of the most state of the art beamforming algorithms used for personal sound zones have been described. The two techniques described was acoustic contrast control (ACC) and pressure matching (PM), where the tradeoff is between the contrast and the reproduction error.

Different state of the art beamforming algorithms have been studied and one is chosen, described and implemented. The algorithm is called Weighted Pressure Matching Total Variation (wPM-TV) and builds on a hybrid between ACC and PM. The TV part of the algorithm is a constraint that ensure an uniform sound image in the dark zone so that the listeners in the dark zone are not disturbed by sudden changes in energy. The simulation showed that algorithm able to create a high contrast between the sound zones and a low reproduction error, where the different parameters, such as the mix between ACC and PM, still needs to be settled for an real implementation. The parameters impact on the performance and results, are described and shown throughout the chapter. Next a description of the different parameters and their impact on the robustness of the algorithm has been made. This is important when implementing it in a real application.

Last a discussion of how to evaluate the results have been made. Both an objective method and perceptual evaluation have been described. Here it was found, that in general 25 dB contrast between the dark and bright zone should be achieved to ensure a general consensus of good separation in personal sound zones [23]. However, this values are highly dependent on the scenario and the signals.

3 System Design and Simulation Environment

In this chapter, the system design and additional simulation will be described. With the only difference being an additional constraint which is made to improve the uniformity and thereby the perceptual experience in the dark zone. It is seen in figure 3.1 that the wPM and wPM-TV algorithm are both very similar in terms of contrast and MSE. However, the real measurement of this additional constraint requires an elaborate listening test. This is not the focus of this project, therefore the less computational complex wPM algorithm is used.

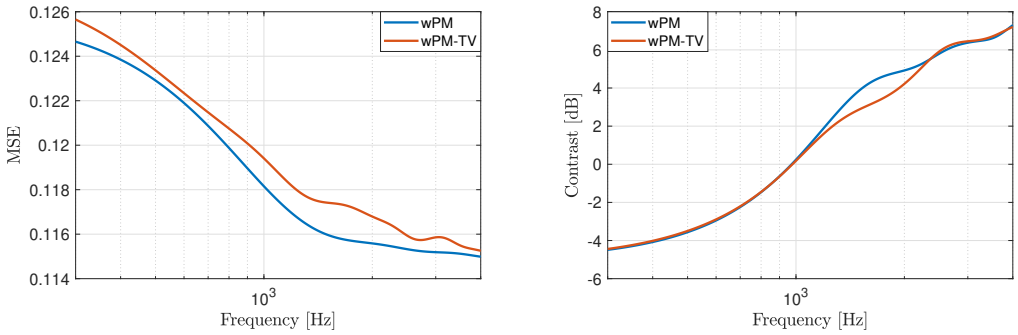


Figure 3.1: Comparison of the contrast and MSE of the wPM and wPM-TV algorithm using 39 loudspeakers in a line array. Bright zone center is placed at [0,2], dark zone center is placed at [0,1], $\lambda = 0.1$, $\xi = 0.9$, $\beta = 0.2$, $\mu = 0.1$, max iterations are 10000, and the stopping criterion is 2^{-23}.

An extension of the algorithm will be made in order to divide the loudspeaker array into smaller subarrays to potentially achieve a higher contrast between a bright and dark zone in sub-optimal positions as described in the introduction. Looking at figure 2.3, the extension will be applied before the algorithm, so that the subarrays will be used as the loudspeakers location input. This system is illustrated in figure 3.2.

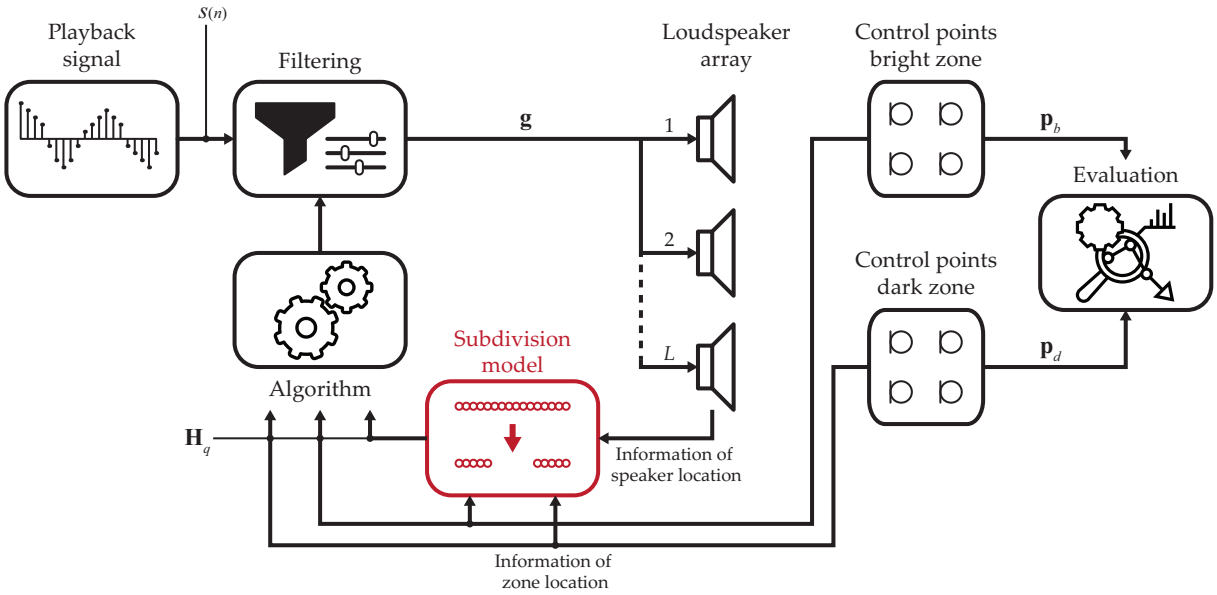


Figure 3.2: Block diagram of the signal path in the system for creating sound zones with a loudspeaker array and with the additional subdivision model.

The method used for dividing the array into subarrays will be made with use of the the ray space transform. It will be investigated how the transform can help by determining which loudspeaker that should be turned off in the different sub-optimal positions. The aim is to define a general method which can then be tested in a real application and compared to the algorithm without the extension.

Before describing the problems with sub-optimal zone positions and describing the ray space transform, it will be examined how the filters from the wPM algorithm behaves in the time domain. It is especially important when the goal is to implement the system in a real application, to ensure that the system is stable e.g. when changing the array effort regularization parameter λ. These test will also help to specify what the different parameters of the algorithm should be.

3.1 Delimitation in Simulation Environment

In this project a line array will be used and placed up against a wall, to ”recreate” a living room scenario with a soundbar under the television. However, the simulation will start by assuming free-field conditions, with loudspeakers modeled as point sources, with a constant speed of sound (343 m/s) and the assumption that the room behaves as an linear-time-invariant (LTI) system. When doing the simulation, no restriction to the amount of loudspeakers used in the array will be made in the simulation. Therefore, different array sizes will be used throughout this chapter to get an idea of the algorithm’s behavior.

The optimization will be done in the frequency domain. This is chosen because it is faster than optimization in the time domain, especially for scaling problems, and Marcos F. Simón Gálvez et al. have in [3] showed almost equal performance as long as the filter length, when using an inverse fast Fourier transform (IFFT), is more than 128 samples. This is not a problem as the focus is not a real time implementation. Analysis of the created impulse response will be done

in order to check for artifacts such as pre-/post-ringing, which can make audible artifacts in the resulting bright zone [5]. To evaluate the results both the MSE and contrast will be used as a measure of performance.

3.2 Test of Weighted Pressure Matching in Time Domain

The most important part of creating the sound zones is the filtering of the signal for each loudspeakers. This can simply be done with a convolution of the signals and the filter impulse response. In this section, a description of how the calculated filters in the frequency domain are transformed into impulse responses in the time domain will be made. The different parameters in the wPM algorithm and their affect on the impulse responses will be tested.

3.2.1 Filters as Impulse Responses

To get the filters impulse responses an inverse fast Fourier transform (IFFT) can be used. However, the amplitude of the impulse and if there is some pre-/post-ringing, needs to be looked at when creating these impulse responses. Pre-/post-ringing can make artifacts in the played sound and aggravate the listening experience [25]. If pre-/post-ringing is a problem, it is possible to apply a filter to the impulse response to shape it and reduce the ringing, without compromising the resulting acoustic separation between the zones [25].

The amplitude of the impulse can provide extra gain in the system. Normally a digital music signal is represented by numbers between -1 and 1 with a given bit-depth [26]. If the numbers becomes larger than ± 1 the signal will overflow and distort. Therefore, to ensure that distortion is not a problem, the amplitude of the impulse response should be within -1 and 1. However, even with the impulses between ± 1, it can still provide more gain in some frequencies resulting in overflow, thus the amplitude response of the filters will also be looked at in the frequency domain.

Creating Filters As Impulse Responses

In the algorithm, the optimization is done in the frequency domain for the frequencies of interest. Just taking the IFFT of the output of the algorithm, will not take the sampling frequency into account. Therefore, some steps before the IFFT is required to create the impulse responses of the filters, the following have been done:

1. Determine which frequencies that should be optimized for. This is determined by a start frequency, f_{start}, a stop frequency f_{stop} and a parameter f_{bins} that controls the resolution. E.g. if $f_{\text{bin}} = 2$, every second frequency between start and stop will be a part of the optimization.
2. The output of the algorithm $\mathbf{g}_{\text{wpm}}$, needs to be zero padded so that the length of the output have a connection to a given sampling frequency, f_s. The amount of zeros that

needs to be added before and after the filter $\mathbf{g}_{\text{wpm}}$ are defined as followed:

$$\text{Before} = \frac{f_{\text{start}}}{f_{\text{stop}}} \Longrightarrow f_{\text{start}} \geq f_{\text{bin}} \tag{3.1}$$

$$\text{After} = \frac{f_s/2}{f_{\text{bin}}} - \frac{f_{\text{stop}}}{f_{\text{bin}}} \tag{3.2}$$

3. The output from step 2 now need a DC component, which should be zero.
4. The new vector from step 3 then needs to be extended with a mirrored version, excluding the DC component and the Nyquist sampling frequency.
5. The IFFT is then used on the output from step 4, and is shifted circularly with half of the length of the output of the IFFT.

In figure 3.3 the impulses of all the filters are shown for a setup where: $\xi = 0.9$, $\lambda = 0.1$, a bright zone is located in [0.5, 2] and the center of the dark zone is placed at [-0.5, 2] with control points placed in a 3×3 with a spacing between each control point of 0.05 m. Filters are calculated from 300 Hz to 3.5 kHz in steps of 1 Hz, and the amount of loudspeakers in the line array are 29 spaced 0.05 m and with center in [0, 0]. The sampling frequency is set to 48 kHz, thus the length of the output of IFFT is 48000 samples when the filter calculations are in 1 Hz steps.

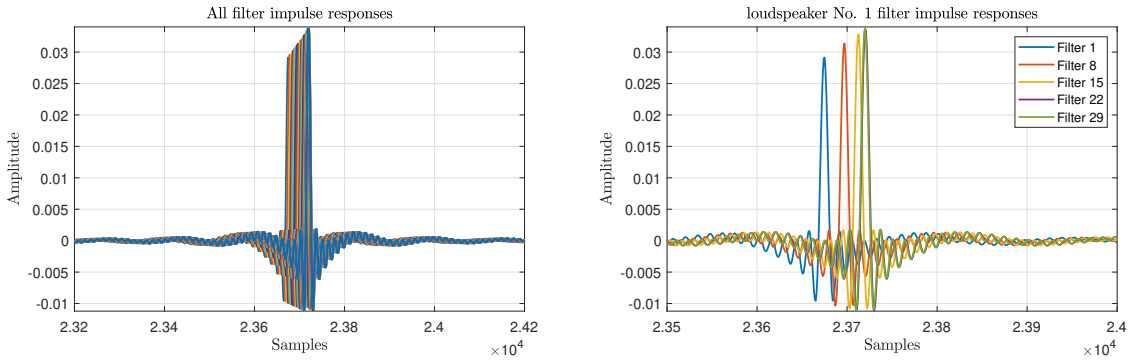


Figure 3.3: The calculated filters for each loudspeaker as impulse responses. The left plot shows all impulses for all loudspeakers, where the right plot shows every 7th impulse zoomed so that the impulses can be seen in more details.

In figure 3.3 it can be seen that ringing before and after the main pulse occurs. This is probably a result of the zero padding of the filter [27]. To test this, the same setup have been tested, but this time, the filter are calculated from 1 Hz to 3.5 kHz, 300 Hz to 3.5 kHz, 300 Hz to 24 kHz (half sampling rate) and 1 Hz to 24 kHz. The results can be seen in figure 3.4.

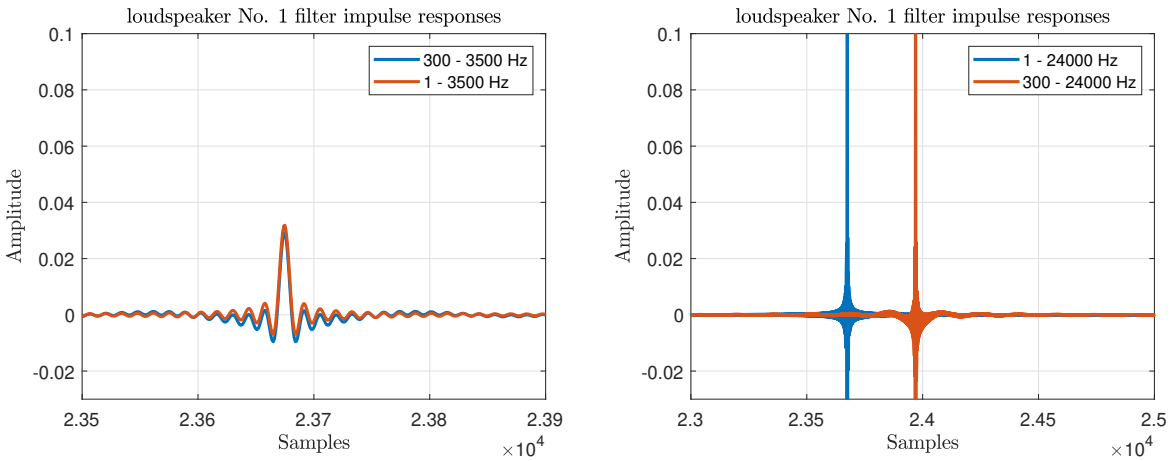


Figure 3.4: Results of using different start and stop frequencies when calculating the filters. The plots shows the effect with and without zero padding.

As it can be seen in figure 3.4, the ringing becomes slightly less when the filters are calculated from 1 Hz to 3.5 kHz. However, the best results are obtained when there is no need for zero padding. Calculating the impulse with less zero padding, also increases the amplitude of the impulse. In figure 3.5 the impulses from figure 3.4 can be seen in the frequency domain from 20 Hz to 24 kHz

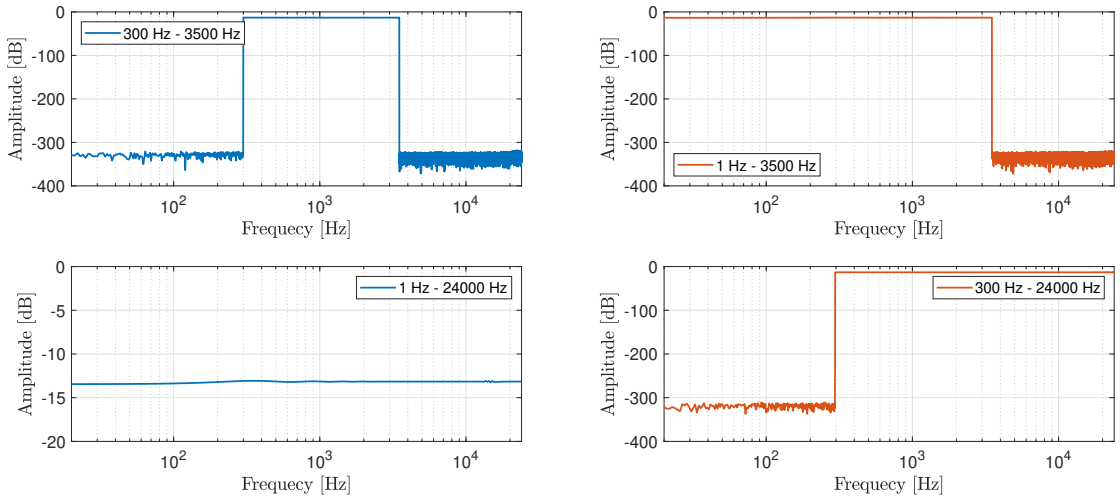


Figure 3.5: Impulse responses from figure 3.4 in frequency domain.

As it can be seen in figure 3.5, sharp edges with noise occurs where the zero padding is done. This is probably the cause of the extra ringing seen in the time domain, thus calculating the filters from 1 Hz to 24 kHz will be preferable.

However, calculating a filter for each frequency can, as mentioned earlier, be very time consuming, as the wPM algorithm optimizes for each frequency. It could therefore be beneficial to calculate the filters with a lower resolution than 1 Hz steps. However, when calculating the filters with lower resolution, less frequencies can be represented. This could be a problem if different frequencies placed close to each other requires different filtering. Thus, if the resolutions

is not high enough the filtering is assumed not to be accurate. Test to see how different frequency steps affects the amplitude response have been made and can be seen in figure 3.6.

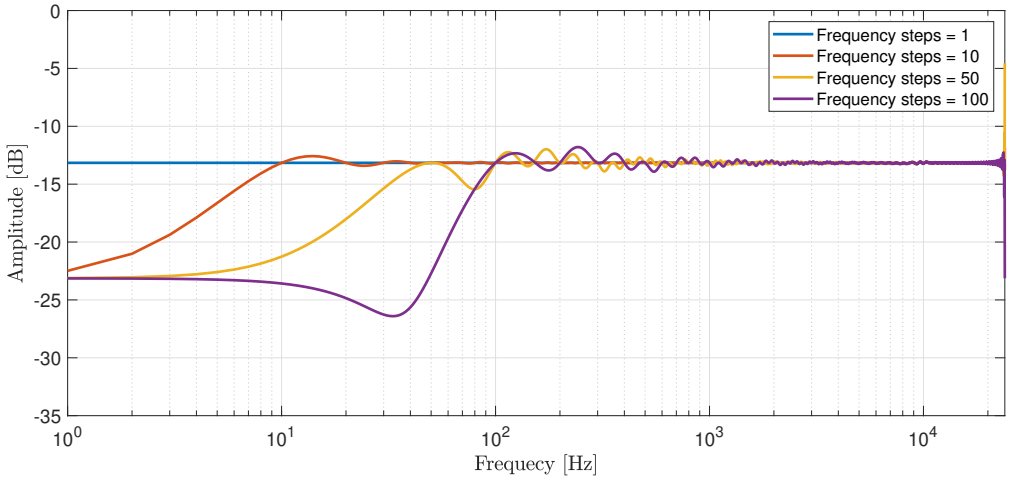


Figure 3.6: Amplitude responses for different impulses calculated with different frequency steps. All the amplitude responses are made with a 48000 points FFT.

It can be seen in figure 3.6 that less frequency bins provide more ripple in the filters compared to the filter calculated with 1 Hz frequency steps. However, as it can be seen, the biggest differences is happening in the low frequencies. Thus might not be a problem when using beamforming, since beamforming works best at mid-range frequencies. To determine the impact that the introduced ripple have on the performance of the beamforming, tests would be required to determine the tradeoff. Therefore frequency steps of 1 Hz will be used for further tests.

To see the regularization parameters, λ, influence on the impulse response, tests with different values of λ have been made. It was found, that λ effects the impulse with more ringing when it becomes smaller. A plot of this can be seen in figure 3.7. The setup is the same as described above, just with different values of λ. Again, for readability, the impulses showed are the ones used for the first loudspeaker in the array beginning from left.

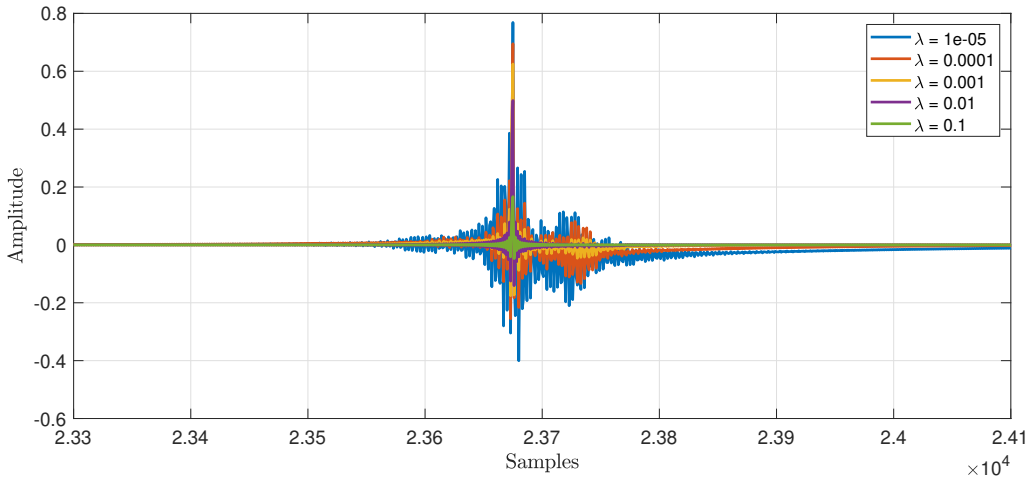


Figure 3.7: Impulse responses created for the same setup but with different values of the regularization parameter λ.

It can be seen in figure 3.7, that the amplitude of the impulse decreases when the value of λ increases. This is probably a result of the decreased array effort in the system with higher values of λ. The increased amount of pre/post-ringing can also be a results of the fact that λ stabilize the matrix inversion, where higher values provide a more stable system as described in section 2.2.1 right under equation (2.14).

To see the amplitude response of the filters and see how they behave at different frequencies, an FFT have been performed on the impulses. The results can be seen in figure 3.8.

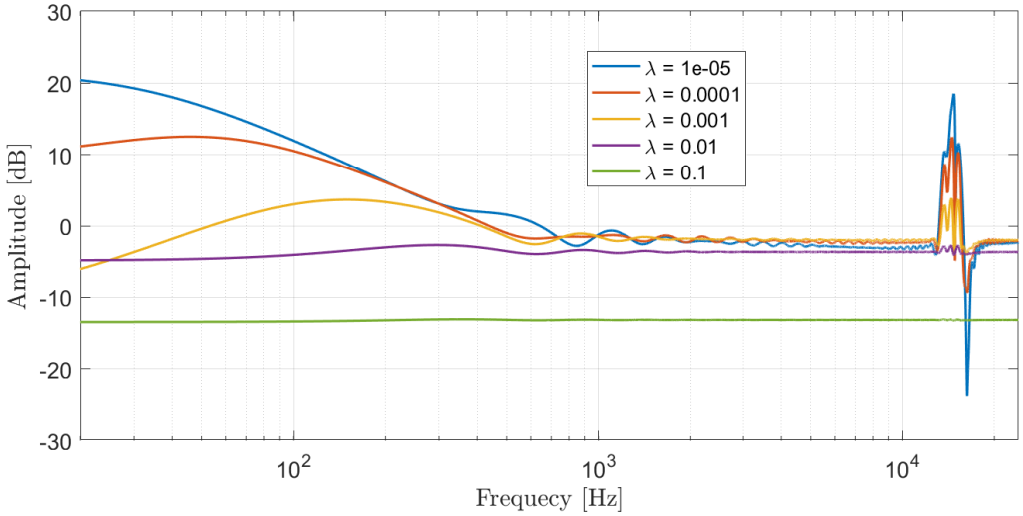


Figure 3.8: Amplitude response for the impulses seen in figure 3.7.

As it can be seen in figure 3.8, when decreasing the value of λ it introduces more ripple in the amplitude response in the high frequencies, and gain in the low frequencies. It can also be seen that all filters calculated with a $\lambda \geq 0.0001$ requires more than 0 dB gain at the lower frequencies. This means that even though the impulses have amplitudes lower than ± 1, there is still a risk that the filtered signal will overflow. E.g. a convolution between a sine wave with frequency 300 Hz and amplitude ± 1, and the impulse for $\lambda = 0.0001$ will result in overflow and therefore a distorted signal.

However, as it can be seen in figure 3.8, the boost in frequencies are outside the range where, based on the size of the array, the beamforming will work best (mid-range frequencies). To see if this holds for all the 29 filters. The amplitude responses of the filters, band pass filtered from 300 Hz to 3.5 kHz, for each value of λ, can be seen in figure 3.9. Additionally, the band pass filtered impulse responses can be seen in figure 3.10.

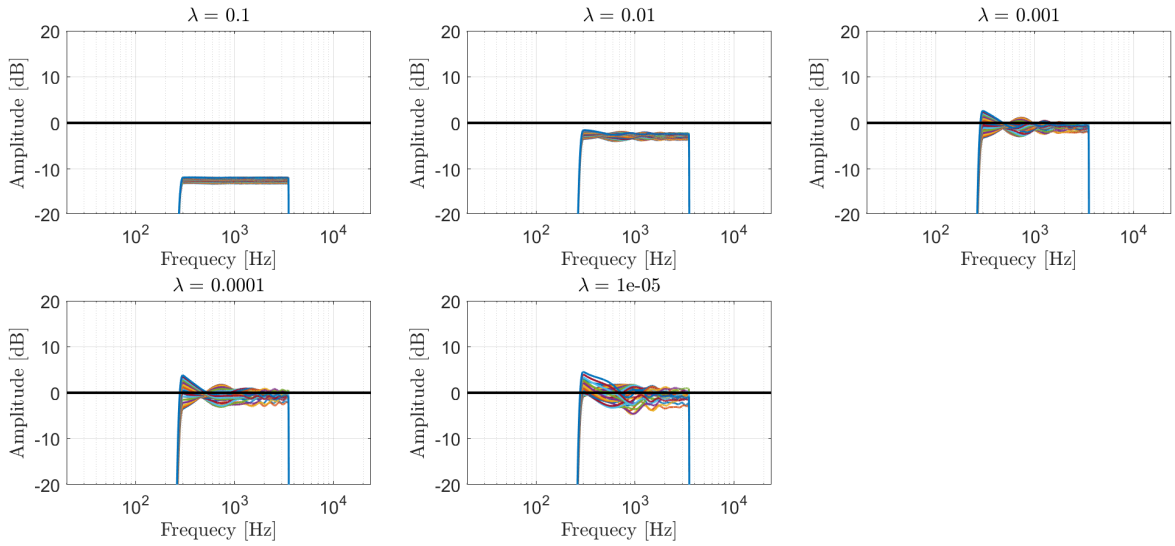


Figure 3.9: All the amplitude responses for the impulses seen in figure 3.7.

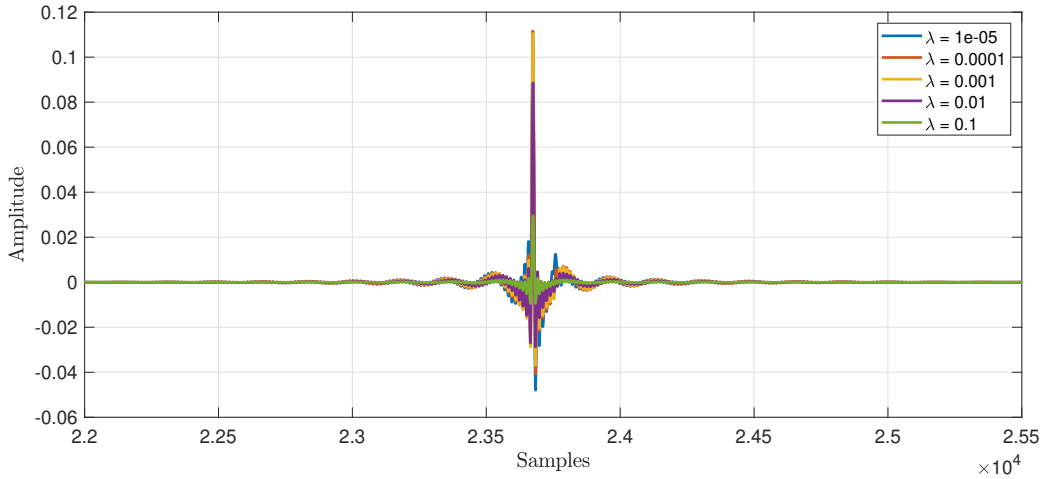


Figure 3.10: Impulse responses created for the same setup but with different values of the regularization parameter λ.

Figure 3.9 shows that some of the filters also boost the signal within the given frequency range (300 Hz to 3.5 kHz). E.g. for $\lambda \leq 0.001$ a boost is applied within the region of interest. However as these filters are very dependent on the specific setup, simulations will be made later to determine the value for λ used for the actual measurements.

As it can be seen in figure 3.10, a lot of the ringing in the filters are removed when using a band pass filter as well as a decrease in the amplitude of the impulses. However, when band pass filtering the filters, ringing in the impulse occurs as seen earlier in figure 3.4.

Impulse Responses and the Number of Loudspeakers

As shown earlier, the regularization parameter λ can provide a higher contrast between the zones, when the value decreases, seen in figure 2.6 in section 2.3. However, this come at a cost of an increased array effort and the smaller the value gets, the less stable the system will be, since λ helps the matrix avoid being ill conditioned, which can cause large numerical errors when doing a matrix inversion [8].

It was found, that when decreasing the number of loudspeakers in the array, while keeping all other settings the same, the filter impulse responses increases in amplitude, with the highest absolute values exceeding 1. This variation was found to be highly dependent on the regularization parameter λ. A plot showing this can be seen in figure 3.11. Here the same zone locations and algorithm settings are used in each test with different amount of loudspeakers and different values of λ. The points show the highest absolute value of all the impulse responses of the filters with a specific amount of loudspeakers. The filters are calculated using the wPM algorithm.

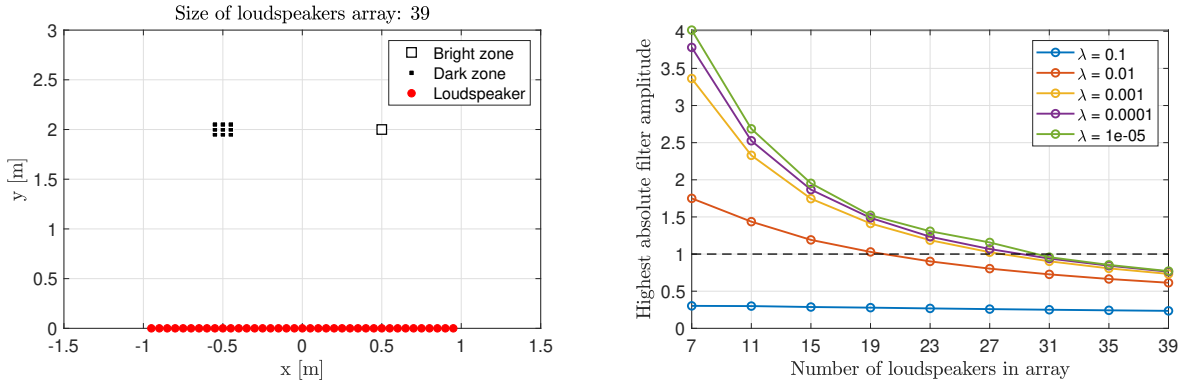


Figure 3.11: Plot showing the impact of the regularization parameter λ on the highest absolute value of all the filter impulse responses with different amount of loudspeakers in the loudspeaker array. The settings in the algorithm are: $\xi = 0.9$, a bright zone located in [0.5, 2] and a 3×3 grid of control points are in the dark zone with center in [-0.5, 2]. Filters are calculated from 1 Hz to 24 kHz in steps of 1 Hz.

It can be seen in figure 3.11 that the highest amplitude of the impulse responses, becomes more constant with a higher value of λ. This test indicates that if loudspeakers are to be turned off, the value of λ is preferred not being too low.

Looking at Zone Impulse Responses

Simulations have been made in order to see how the impulses look in the different control points. To get the impulse from a point, an estimated transfer function from a loudspeaker to a point, is made using equation (2.4). Each distance from a loudspeaker in the array to the point is calculated and used in (2.4), where the filters calculated for the loudspeakers are multiplied with the transfer functions in the frequency domain.

As the sound pressure is only optimized to the exact position of the control points, it could be interesting to see how big of a difference it makes not to be in the exact spot. Therefore points surround the control point have been examined. These results can be seen in figure 3.12, where

different values of λ also are included. The filters are calculated from 1 Hz to 24 kHz in 1 Hz frequency steps. Based on the size of the array, the results have been bandpass filtered from 300 Hz to 3.5 kHz. Note that this will decrease the amplitude of the impulse.

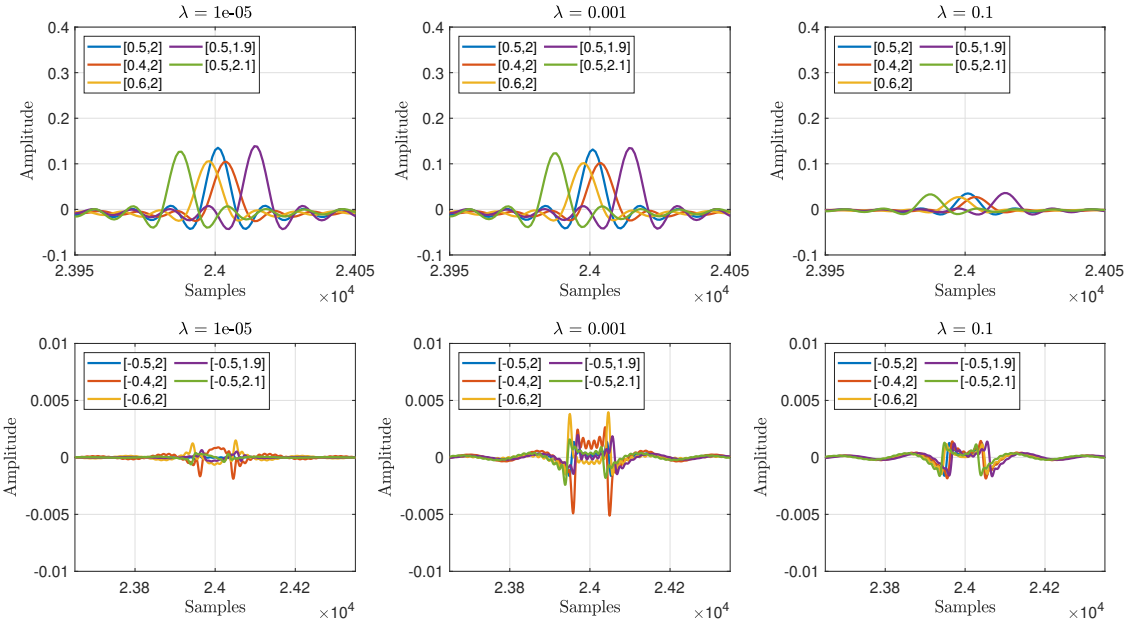


Figure 3.12: The impulses in points surrounding the control point in the bright zone can be seen in the top row and the impulses in points surrounding the center of control points in the dark zone, can be seen in the bottom row.

An additional plot have been made, showing the highest value of the impulse as a function of movement in different directions from the control point. Two figures are shown, one with movement in the x-direction, with the y-coordinate fixed to 2, and the other with movement in the y-direction with the x-coordinate fixed to 0.5, this is shown for different values of λ and can be seen in figure 3.13.

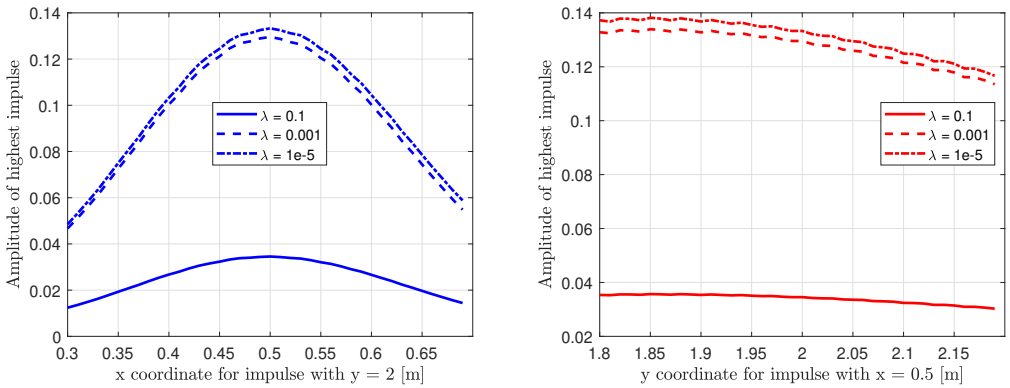


Figure 3.13: Results of impulses in points close to the control point in the bright zone in different directions, showing the highest amplitude of impulse response.

As it can be seen in figure 3.12 and 3.13, the biggest changes in the amplitude of the impulse happens when moving away from the control point in the x-direction and only small changes in the y-direction. This is expected as moving outside of the direction of the sound, will reduce the impulse more than moving in the direction of the sound.

3.2.2 Summary

In this section, calculated filters have been transformed into the time domain, where different parameters, that could have an impact, have been tested. It was found that optimizing from 1 Hz to 24 kHz corresponding to half the sampling rate gave the least ringing in the impulse response. When decreasing the value of λ more ringing in the filters were introduced and also higher gain at the lower frequencies and at frequencies higher that 10 kHz. It was also seen that when decreasing the amount of loudspeakers used, the highest amplitude of the filter impulse responses increased when using values of λ below 0.1. Thus, $\lambda = 0.1$ will be used for further tests in this chapter.

3.3 Sub-Optimal Sound Zone Positions

As described in the introduction, sub-optimal positions of the individual zones might be an issue when trying to achieve high contrast between the two sound zones. The simulations showed so far have been for zones placed without shadowing for each other. However simulations have been made in order to show how the algorithm perform in sub-optimal positions. The contrast is calculated as the mean difference in dB between the control points in the two zones. In addition, simulations where some of the loudspeakers are turned off have been made. This is done in order to see how strategically removing some of the loudspeakers can affect the contrast and the MSE. This can be seen in figure 3.14, where what is assumed to be the least optimal positions are simulated. The simulation illustrates two listeners standing in a line in front of each other. The simulation shows the contrast from 300 Hz to 3.5 kHz (in steps of 1 Hz), which consists of 89 point sources placed 0.05 m from each other, this results in an array width of 4.4 m. The surface plots are shown for 1 kHz. The bright zone is a single control point placed in [0, 2] and the dark zone is a 3×3 grid of control points spaced 0.05 m from each other with center in [0, 1]. The weighting factor and regularization parameter are chosen to $\xi = 0.9$ and $\lambda = 0.1$. In figure 3.15, the sound zones have been swapped.

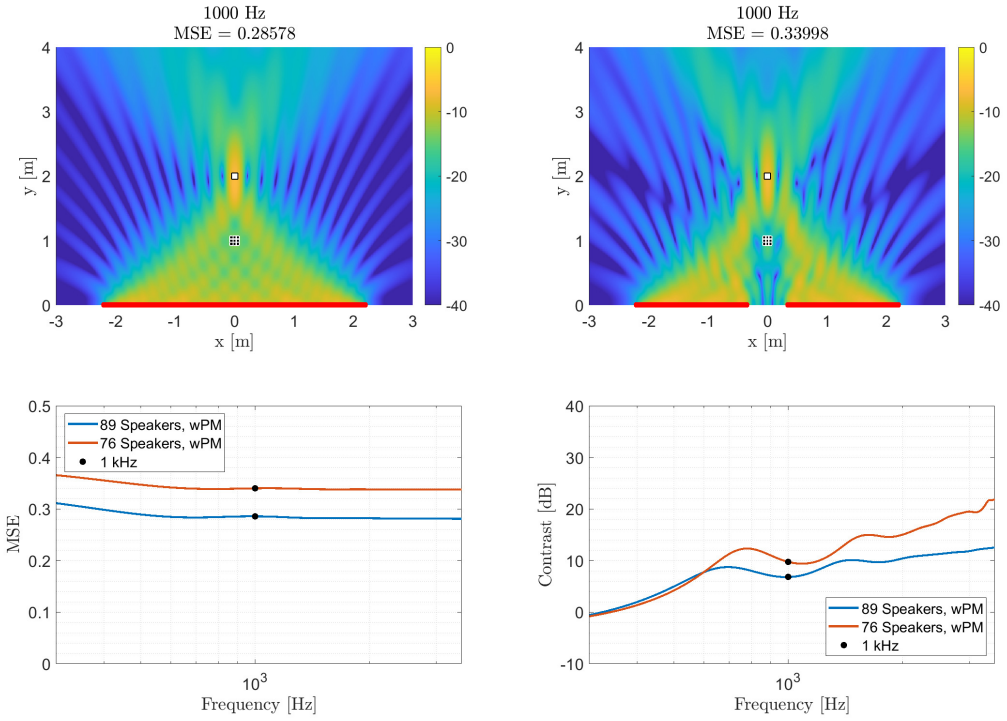


Figure 3.14: Simulation of a sub-optimal position with and without loudspeakers removed shown for 1 kHz, where the red dots are the loudspeakers. The contrast is shown for frequencies from 300 to 3500 Hz, in steps of 1 Hz, and the contrast is given as the mean difference in dB between the dark and bright zone. Blue line is the contrast before removing loudspeakers and the red line is after.

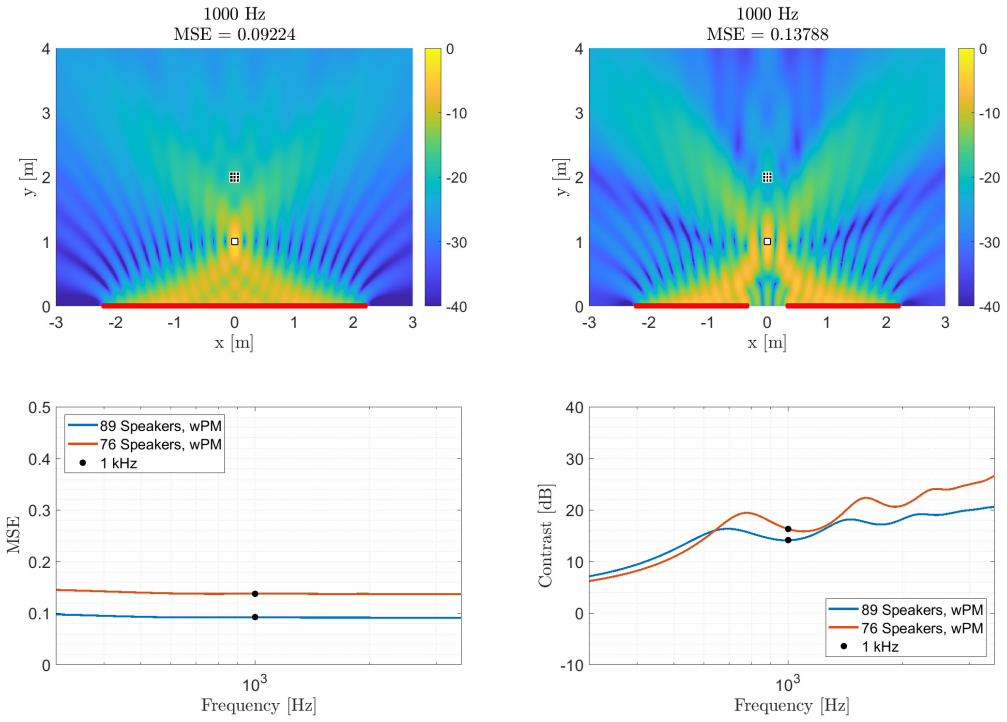


Figure 3.15: Simulation of a sub-optimal position with and without loudspeakers removed shown for 1 kHz, where the red dots are the loudspeakers. The contrast is shown for frequencies from 300 to 3500 Hz, in steps of 1 Hz, and the contrast is given as the mean difference in dB between the dark and bright zone. Blue line is the contrast before removing loudspeakers and the red line is after.

As it can be seen, both in figures 3.14 and 3.15, the contrast is increased when turning off loudspeakers, where a lot of the sound have to travel directly through the one sound zone to get to the other. This indicates that when only using parts of the array to create the sound zones in sub-optimal positions, higher levels of contrast can be achieved. However this comes with an increased MSE.

To see another sub-optimal position, which is not as critical as the first one, additional simulations have been done. All of the settings are the same as in figure 3.14, but the bright zone is now placed in [-0.5, 2] and the dark zone is placed with center in [0.5, 1]. The results can be seen in figure 3.16. Where in figure 3.17, the zones have been swapped.

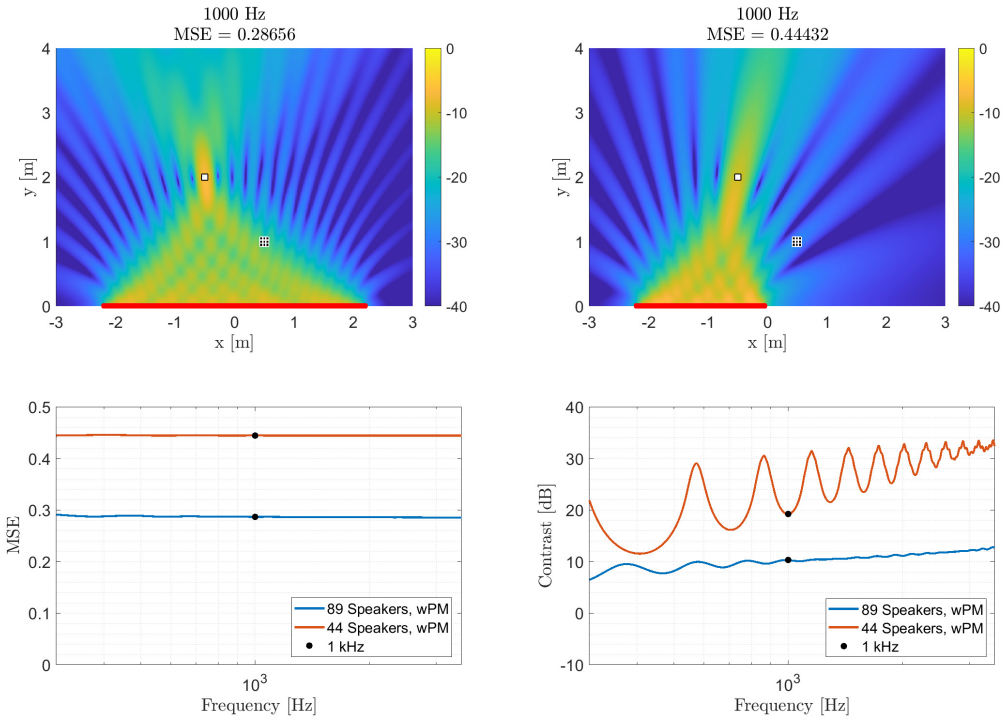


Figure 3.16: Simulation of a sub-optimal position with and without loudspeakers removed shown for 1 kHz, where the red dots are the loudspeakers. The contrast is shown for frequencies from 300 to 3500 Hz, in steps of 1 Hz, and the contrast is given as the mean difference in dB between the dark and bright zone. Blue line is the contrast before removing loudspeakers and the red line is after.

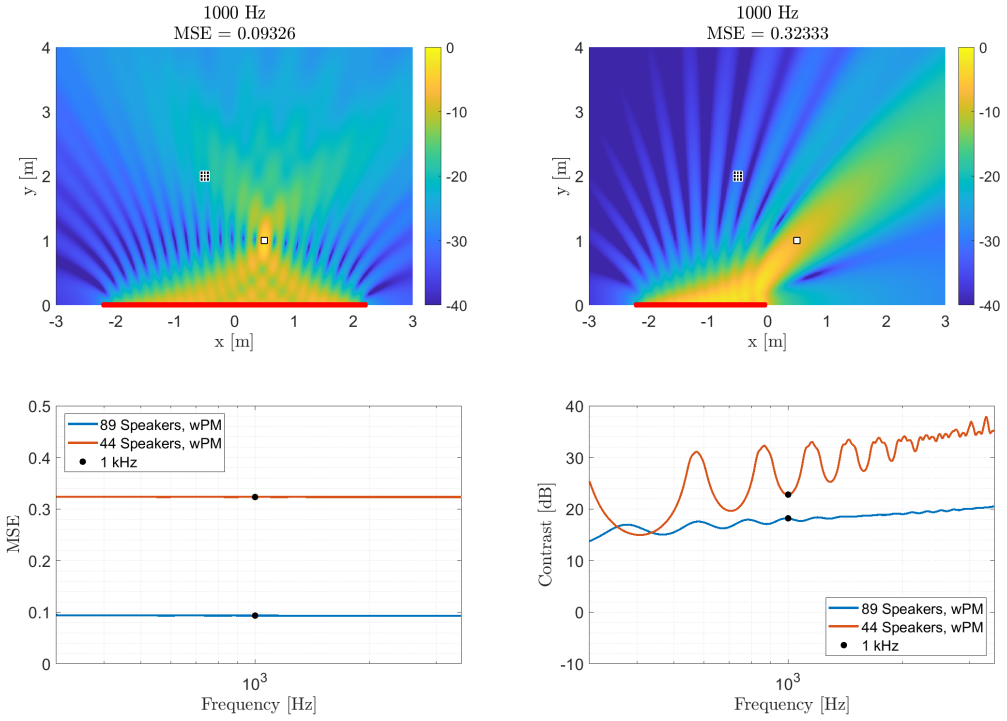


Figure 3.17: Simulation of a sub-optimal position with and without loudspeakers removed shown for 1 kHz, where the red dots are the loudspeakers. The contrast is shown for frequencies from 300 to 3500 Hz, in steps of 1 Hz, and the contrast is given as the mean difference in dB between the dark and bright zone. Blue line is the contrast before removing loudspeakers and the red line is after.

Again a high increase in the contrast can be seen when turning off loudspeakers, along with an increased MSE as it also was seen in figure 3.14 and 3.15.

This indicates, that an extension to the algorithm, to tell whether one zone is shadowing for the other and which of the loudspeakers that should preferable be turned off, could be use. However, removing loudspeakers showed an increase in the MSE. Meaning that reproduced sound is less accurate, which can be due to less loudspeakers recreating the desired the sound field.

3.4 Subarray Division by the Use of the Ray Space Transform

In the sub-optimal setups as described in the introduction and showed in the former section, there are cases where the one zone are blocking some of the loudspeakers direct path to the other zone. In such cases it might be beneficial to turn off these loudspeakers in order to achieve a higher contrast between the zones, this was shown in section 3.3. In order to identify these loudspeakers, the ray space transform will be used as an analysis tool. The ray space is a domain where directional information and components of the acoustic field are mapped onto, where relevant acoustic object becomes linear patterns [28]. In earlier studies, the ray space transform have been primarily used on microphone arrays, where one of the advantages, when using the ray space transform, is, that it is possible to assume far-field in near-field situations because of the use of subarrays [29, 28]. To get a better understanding of how the ray space transform works, it will first be described by its original use. Afterwards, it will be explained how the ray space domain can be used in order to divide the loudspeaker array into smaller subarrays.

Before applying the ray space transform a reference frame is made in the geometric domain. The geometric domain can be seen as the "real application", where the acoustic events happen. An illustration of such acoustic scene can be seen in figure 3.18.

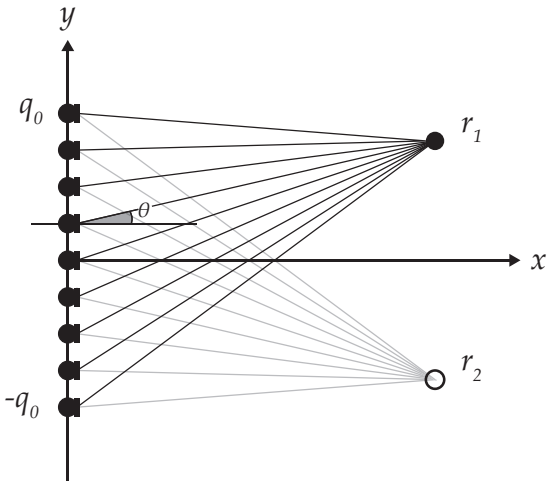


Figure 3.18: Illustration of the geometric domain, where two point sources are placed directly next to each other, seen from the microphone array.

The microphone array lies on the y-axis between $y = -q_0$ and $y = q_0$, where this area also determine the size of the observation window (OW). In 2D geometry, the OW describes a line segment through which the acoustic scenes are "observed" [30]. The point r_1 and r_2 are the two different point sources. The ray space transform is then given as the linear equation:

$$y = mx + q \tag{3.3}$$

where the parameters $(m = \tan(\theta), q)$ describes all lines that is not parallel to the y-axis [28, 30]. The space of such parameters is called the ray space and an illustration going from the geometric domain to the ray space domain can be seen in figure 3.19.

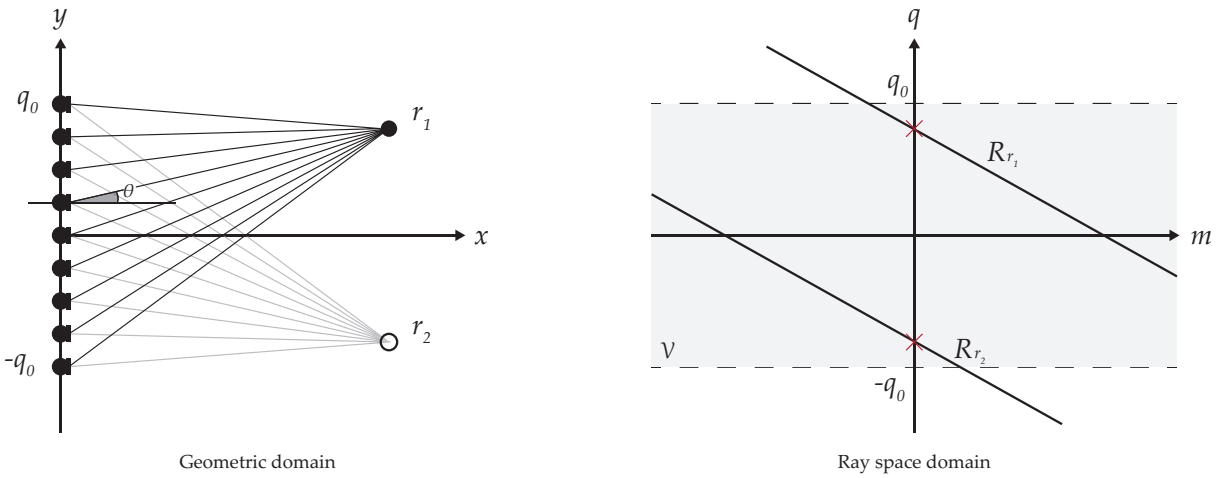


Figure 3.19: Illustration of the geometric domain mapped onto the ray space.

The lines seen in ray space domain in figure 3.19 are assumed to be coming from the positive half-space of the geometric domain where $x > 0$ [28], which is also where the sound field of interest (also referred to as "region of interest" (ROI)) is. The grey area the ray space is called the "visibility region". Given a space $\mathcal{P}$ of all possible parameters (m, q) in the positive half-space, the visibility region can be written for lines that only are within the OW as [30]:

$$\mathcal{V} = \{(m, q) \in \mathcal{P} : -q_0 \leq q \leq q_0\} \tag{3.4}$$

In order to understand the structure of the ray space, a point $r = [x, y], x > 0$, represented in the ray space, is considered. This point can equivalently be thought as a set of all the lines r that pass through it. Where these lines only identify those rays that depart from the source and point towards the y-axis [30]. In the ray space, the region of parameters describing such lines can be represented as [28, 30]:

$$q = -xm + y \tag{3.5}$$

The ROI of r is then the set of lines that passes through both r and the OW [30], which can be described as:

$$\mathcal{R}_r = \{(m, q) \in \mathcal{V} : q = -xm + y\} \tag{3.6}$$

These lines can also be seen in the ray space domain shown in figure 3.19. It can be seen that the slope of $\mathcal{R}_r$, is determined by the x-coordinate of the point source, meaning that the steepness of the slope is directly correlated to how far the point source is from the microphone array. The y-coordinate determine where on the q-axis in the ray space domain that the line $\mathcal{R}_r$ intersect. An illustration of different point source positions and the different steepness of the slopes in the ray space domain can be seen in figure 3.20.

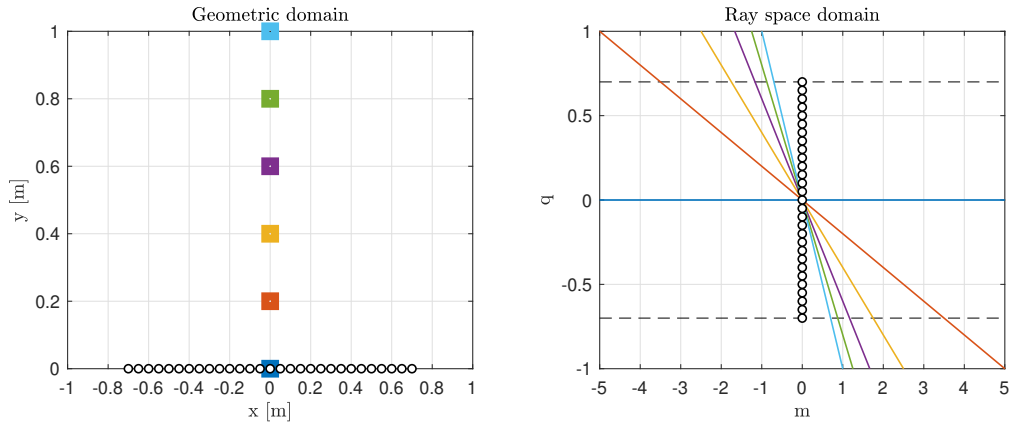


Figure 3.20: Plot showing the connection between the distance from the microphone array to a point source in the geometric domain and the slope of the associated line in the ray space domain.

However the application for this project, is with an array of loudspeakers instead of microphones and with individual control points instead of point sources. This means that the ROI which was for the microphone array seen as the OW, is instead the area in which loudspeakers are placed. If the intersections between the different control points in the ray space domain are placed within the ROI. This means that the direct path between a control point and somewhere on the loudspeaker array is blocked by another control point.

However when using control points and a loudspeaker array instead of point sources with a microphone array, the ray space transformation will still be the same, thus the line $\mathcal{R}_{r_1} = \mathcal{R}_{p_B}$ for the bright zone and the line $\mathcal{R}_{r_2} = \mathcal{R}_{p_D}$ for the dark zone, this can be seen in figure 3.21.

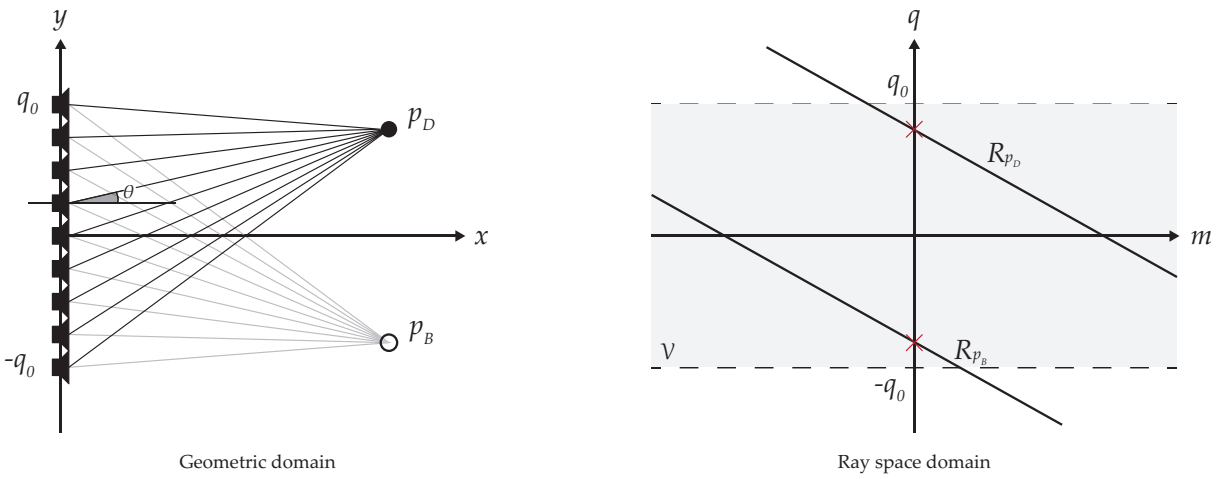


Figure 3.21: Illustration of the geometric domain mapped onto the ray space for a loudspeaker array and two control points.

3.4.1 Dividing the Loudspeaker Array

As described earlier the position of the control points in the geometric domain, is preserved into the ray space domain in the steepness of the slope and the intersection of the lines and the q-axis.

The ray space transform has been implemented in MATLAB, going from the geometric domain to the ray space domain. Where the bright and dark zone coordinates are vectors with the coordinate of every control point. Additionally the intersections between lines from the bright zone control point and the dark zone control points, $\mathcal{R}_{p_B}$ and $\mathcal{R}_{p_D}$, if there is one, have been calculated as:

$$\mathrm{I}_m = \frac{\mathcal{R}_{B,y} - \mathcal{R}_{D,y}}{\mathcal{R}_{B,x} - \mathcal{R}_{D,x}} \tag{3.7}$$

$$\mathrm{I}_q = -\mathcal{R}_{B,x} \cdot \mathrm{I}_m \ + \mathcal{R}_{D,y} \tag{3.8}$$

Where $I_{m,q}$ is the set of intersection points between the lines from the dark zone control points and the bright zone control point, this is described as:

$$I_{m,q} = \{(m,q) : (m,q) \in \mathcal{R}_{p_B} \cap \mathcal{R}_{p_D}\} \tag{3.9}$$

The MATLAB implementation of the ray space transform can be seen in code example 3.1.

Code example 3.1: Implementation of ray space transform in MATLAB.

```
q = [BrightZoneCoordinates(:,1)-m.*BrightZoneCoordinates(:,2);...
     DarkZoneCoordinates(:,1)-m.*DarkZoneCoordinates(:,2)];
     %X and Y coordinates swapped to get the loudspeakers on the q-axis in the ray space domain.

crossM = (DarkZoneCoordinates(:,1)-BrightZoneCoordinates(:,1)')./...
          (DarkZoneCoordinates(:,2)-BrightZoneCoordinates(:,2)');
crossQ = BrightZoneCoordinates(:,1)'-crossX.*BrightZoneCoordinates(:,2)';
CrossCoordinates = [crossM,crossQ]; %set of coordinates of where the lines cross in the ray
    space.

LinesY = q'; %Matrix with q-values of the lines in the ray space
LinesX = m'*ones(1,length(q(:,1))); % %Matrix of m-values of the lines in the ray space
```

It can be seen in figure 3.21, that the lines, $\mathcal{R}_{p_B}$ and $\mathcal{R}_{p_D}$, in the ray space ray space domain are parallel. This means that the control points are placed directly next to each other in the geometric domain seen from the loudspeaker array and that none of the control points are blocking the direct path from a loudspeaker to a point. As described earlier intersections in the ROI means that some of the control points are in front of each other. This can be seen with different control point placements in the figures 3.22, 3.23 and 3.24.

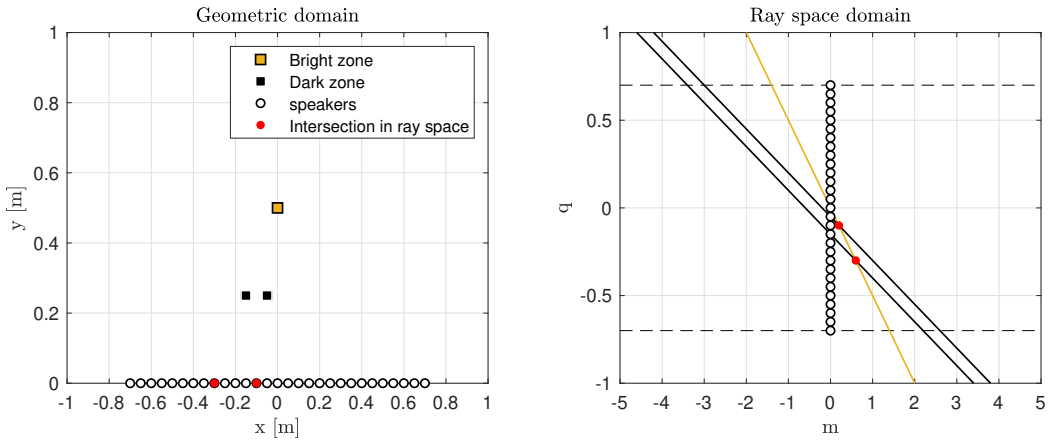


Figure 3.22: Plot of geometric domain and its equivalent in the ray space domain. A bright zone located in [0, 0.5] and a dark zone with two control points with center in [-0.1, 0.25], with 0.1 m spacing between the points. The loudspeaker array consist of 29 loudspeakers center in [0, 0] with 0.05 m displacement.

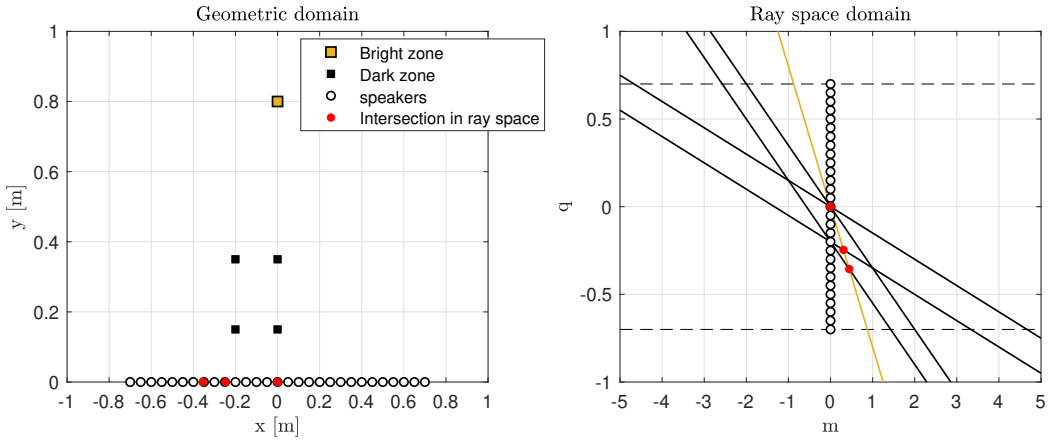


Figure 3.23: Plot of geometric domain and its equivalent in the ray space domain. A bright zone located in [0, 0.8] and a dark zone with a grid of 2×2 control points centered in [-0.1, 0.25], with 0.2 m spacing between the points. The loudspeaker array consist of 29 loudspeakers center in [0, 0] with 0.05 m displacement.

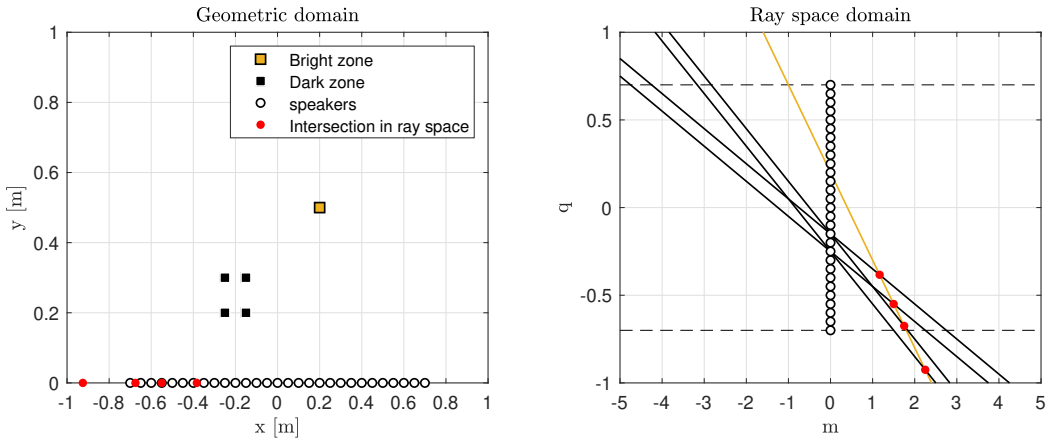


Figure 3.24: Plot of geometric domain and its equivalent in the ray space domain. A bright zone located in [0.2, 0.5] and a dark zone with a grid of 2×2 control points centered in [-0.2, 0.25], with 0.1 m spacing between the points. The loudspeaker array consist of 29 loudspeakers center in [0, 0] with 0.05 m displacement.

Figures 3.22, 3.23 and 3.24, all show the concept of the ray space transform in different scenarios. The q value can be used in the geometric domain to tell, where on the loudspeaker array these intersections are happening. If the point is within the ROI and is in the same place as a loudspeaker, it should probably be turned off. But as it can be seen in figure 3.24, two of the points within the ROI are not directly hitting a loudspeaker, thus it should be determined whether a loudspeaker (or maybe several) should be turned off.

The relative position of the sound zones can be determined by looking at where the intersection points are located in the ray space domain and the relative steepness of the slopes. In figure 3.25, a similar setup as figure 3.24, can be seen, with the only difference that the placement of the zones have been mirrored.

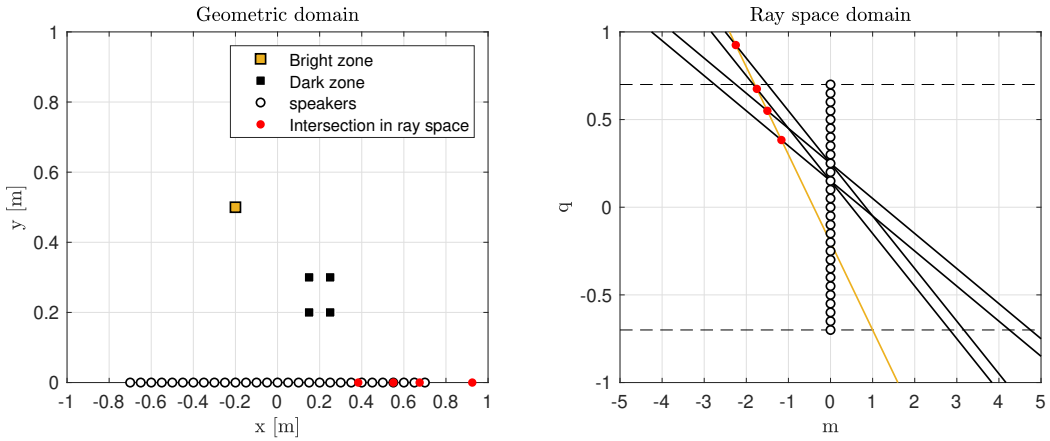


Figure 3.25: Plot of geometric domain and its equivalent in the ray space domain. A bright zone located in [-0.2, 0.5] and a dark zone with a grid of 2×2 control points centered in [0.2, 0.25], with 0.1 m spacing between the points. The loudspeaker array consist of 29 loudspeakers center in [0, 0] with 0.05 m displacement.

It can be seen that in the ray space domain, all the intersection points have been moved to the left half-plane, compared to figure 3.24. With the bright zone slope being steeper than the dark zone slopes, the means that the dark zones are placed closer to the loudspeaker array, and with all of the intersection points placed on the left half-plane, this means that the bright zone is placed to the left of the dark zone. However, if the zones were to switch positions, the slope of the dark zone would be steeper than the bright zone, and the intersection points would be moved to the right half-plane. Meaning that the bright zone would be to the right of the dark zone. This is seen in figure 3.26.

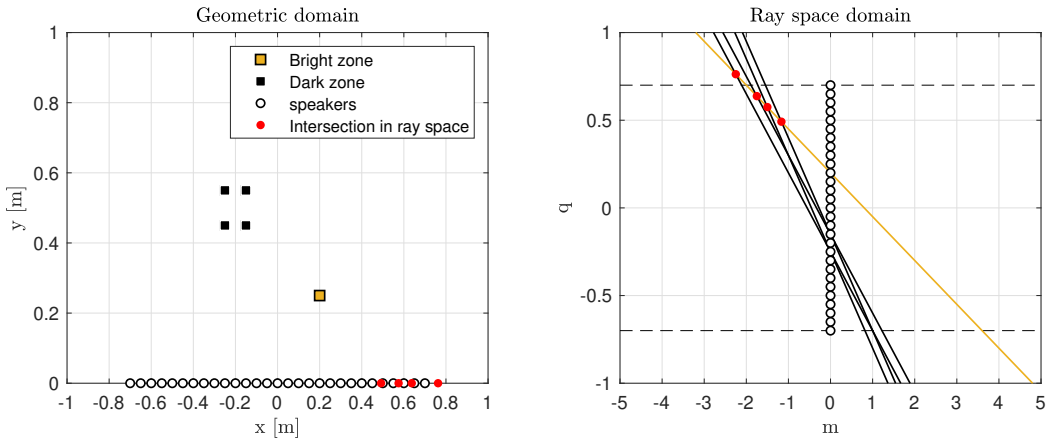


Figure 3.26: Plot of geometric domain and its equivalent in the ray space domain. A bright zone located in [0.2, 0.25] and a dark zone with a grid of 2×2 control points centered in [-0.2, 0.5], with 0.1 m spacing between the points. The loudspeaker array consist of 29 loudspeakers center in [0, 0] with 0.05 m displacement.

3.4.2 Implementing Ray Space Transform Into the System

To help determine whether a loudspeaker should or should not be turned off, the relative position of the zones can be relevant since the choice might not only be limited to the span between the intersection points in the ray space domain. It might be beneficial to turn off the entire side of the array. With the ray space transform implemented the next step is to implement it such that the system can divide the loudspeaker array into smaller subarrays by itself. The principle of this model, is based on the direct correlation between the q-coordinate of the intersection points, $I_{m,q}$, in the ray space domain with and the equivalent x-coordinate in the geometric domain. Two types of subdivision models have been made. The simplest one being the ray space middle subdivision model (RSMS). In this model, the concept is that the intersect points, in the geometric domain are placed on the array-axis. Starting with a set of loudspeaker positions P_{Spk}, the new set of loudspeaker positions, $P_{\text{Spk,RSMS}}$, where the loudspeakers placed in between intersection points have been turned off, can be described as in equation 3.10 and can be seen implemented in code example 3.2:

$$P_{\text{Spk,RSMS}} = \{x \in P_{\text{Spk}} : I_{q,\max} < x \ \cup \ I_{q,\min} > x\} \tag{3.10}$$

Code example 3.2: Implementation of the ray space middle subdivision model in MATLAB.

```
[CrossMax, IdxMax] = max(CrossCoordinates(:,1));
[CrossMin, IdxMin] = min(CrossCoordinates(:,1));


if strcmp(Form,'middle')
    SpeakerVector = find(SpeakerCoordinates(:,1) < CrossCoordinates(IdxMax,2) | ...
                         SpeakerCoordinates(:,1) > CrossCoordinates(IdxMin,2));
    SpeakerCoordinatesNew = SpeakerCoordinates(SpeakerVector,:);
end
```

The other subdivision model, is the ray space edge subdivision model (RSES). In this model, the relative position of the sound zones is taken into account. In cases where the one sound zone is placed to the right and the other to the left, the RSES-model, turns off the entire side of the loudspeaker array. It was seen that if the bright zone was placed to the left of the dark zone, the intersection points was placed in the negative half-plane, and in the positive half-place if placed to the right. This is used in order to determine which of the loudspeakers that should be turned off. The RSES-model finds the intersection point closest to the center of the loudspeaker array, and remove every loudspeaker between this point and the edge of the array opposite to the bright zone. If the intersection points are both in the positive and negative half-plane, the subdivision will be identical to the RSMS-model. This can be described as:

$$P_{\text{Spk,RSES}} = \begin{cases} \{x \in P_{\text{Spk}} : I_{q,\min} < x\} & \text{If} \quad I_{q,\min} > 0 \wedge I_{q,\max} > 0 \\ \{x \in P_{\text{Spk}} : I_{q,\max} > x\} & \text{If} \quad I_{q,max} < 0 \wedge I_{q,\min} < 0 \\ \{x \in P_{\text{Spk}} : I_{q,\max} < x \ \cup \ I_{q,\min} > x\} & \text{If} \quad I_{q,\max} \geq 0 \wedge I_{q,\min} \leq 0 \end{cases} \tag{3.11}$$

Where $P_{\text{Spk,RSES}}$ is the new set of loudspeakers coordinates. This can be seen in code example 3.3.

Code example 3.3: Implementation of the ray space egde subdivision model in MATLAB.

```
[CrossMax, IdxMax] = max(CrossCoordinates(:,1));
[CrossMin, IdxMin] = min(CrossCoordinates(:,1));

if strcmp(Form, 'edge')
    if CrossMin > 0
        SpeakerVector = find(SpeakerCoordinates(:,1) > CrossCoordinates(IdxMin,2));

        SpeakerCoordinatesNew = SpeakerCoordinates(SpeakerVector,:);
    elseif CrossMax < 0
        SpeakerVector = find(SpeakerCoordinates(:,1) < CrossCoordinates(IdxMax,2));
        SpeakerCoordinatesNew = SpeakerCoordinates(SpeakerVector,:);
    else
        SpeakerVector = find(SpeakerCoordinates(:,1) < CrossCoordinates(IdxMax,2) |
            SpeakerCoordinates(:,1) > CrossCoordinates(IdxMin,2));
        SpeakerCoordinatesNew = SpeakerCoordinates(SpeakerVector,:);
    end
end
```

Examples of RSES and RSMS can be seen in figure 3.27. Showing a situation where turning off the entire side of the array provides a better contrast at the cost of a higher MSE.

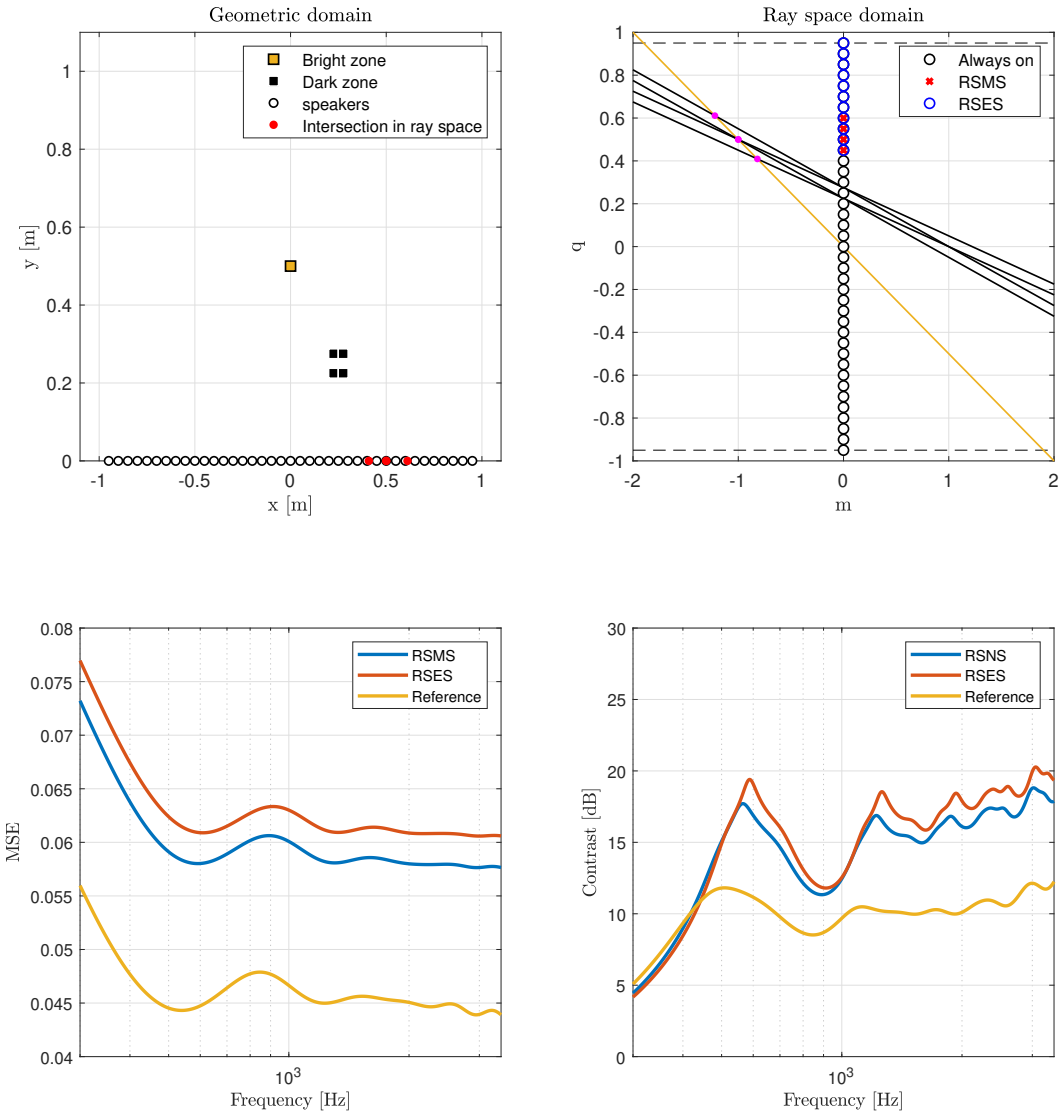


Figure 3.27: Plot of geometric domain and its equivalent in the ray space domain. A bright zone located in [0, 0.4] and a dark zone with a grid of 2×2 control points centered in [-0.2, 0.2], with 0.1 m spacing between the points. The loudspeaker array consist of 29 loudspeakers center in [0, 0] with 0.05 m displacement.

3.5 Test of Ray Space Subdivision of the Loudspeaker Array

In order to test the subdivision models, several simulations have been made. In these simulations, both the RSMS-model and the RSES-model explained in 3.4.2 have been used. The MATLAB implementation of the subdivision models finds the indexes of the loudspeaker placement vector and compares the values to the intersection points in order to decide which of the loudspeakers that should be turned off. However in this section an extension have been made, in order to check if removing more or less loudspeakers than defined by the subdivision models, does improve the results. This is achieved by altering both the start index and stop index by a specified amount, which in this case is ± 1, this alteration is specified with a number after the model name. For example, RSMS1 denotes that the ray space middle subdivision model is used and that 1 additional loudspeaker is removed from each of the two subarrays created by the RSMS-model. This is illustrated in figures 3.28 and 3.29. The first setup illustrated is with the center of the bright zone in $[-0.2, 2]$ and the center of the dark zone in $[0.2, 1]$ with the control points placed in a 3×3 grid, spaced 0.05 m between each control point, the results can be seen in figure 3.28. The second setup is with bright zone control point in $[0, 2]$ and the center of the dark zone in $[0, 1]$ with the control points placed in a 3×3 grid spaced 0.05 m between each control point. For both setups the algorithm settings were chosen to $\xi = 0.9$ and $\lambda = 0.1$.

The results shown that removing additional loudspeakers provide a higher contrast as well as an increased MSE. There seems to be a correlation between the amount of loudspeakers turned off and the contrast achieved and the MSE. This allows a more specific application where the use of the subdivision models can be tuned by increasing or decreasing the amount of loudspeakers turned off. In general it seems that the RSES-model is performing best with contrast but the worst with MSE. This is probably due to the increased amount of loudspeakers removed in this model compared to the RSMS-model, which also shows that the MSE and contrast is correlated with the amount of loudspeakers turned off.

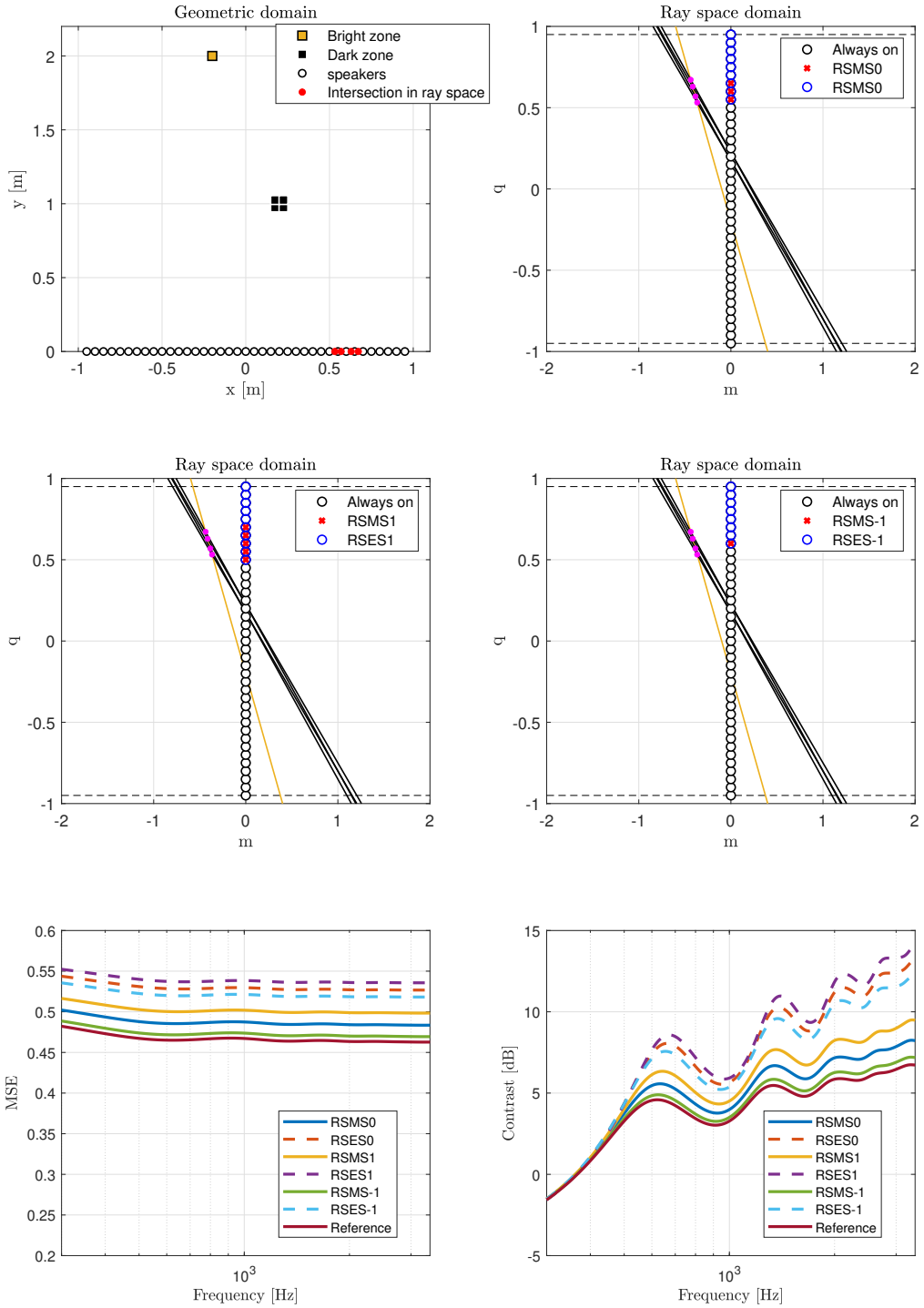


Figure 3.28: Contrast and MSE of different loudspeaker settings, determined by the use of the ray space subdivision models. The different colors of the loudspeakers mark which of them are turned off with a given subdivision model.

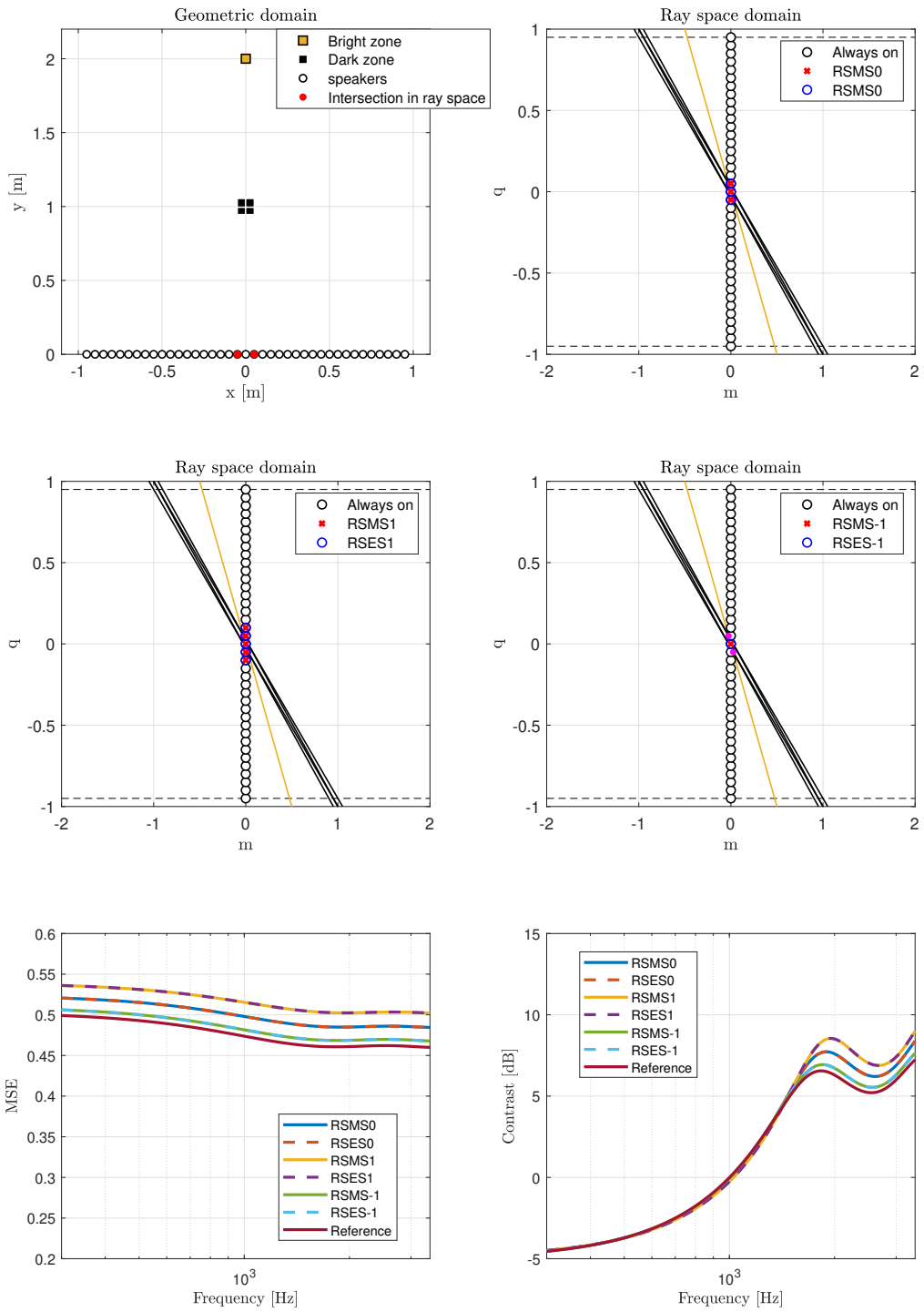


Figure 3.29: Contrast and MSE of different loudspeaker settings, determined by the use of the ray space subdivision models. The different colors of the loudspeakers mark which of them are turned off with a given subdivision model.

4 | Implementation and Physical Limitations of Multizone System

After having simulated different scenarios, proposed a system design and divided the line array into subarrays dependent on sub-optimal positions, the practical implementation, and limitations, of such system will be described in this chapter. The simulations done so far have been with a simple point source model, where in this chapter, two additional models of the acoustic transfer function (ATF) will be introduced: a piston model and a room reflection model using the image source method (ISM). Afterwards, the measurement setup will be described.

4.1 Piston Model and Loudspeaker Radiation

When using loudspeakers there is some physical limitations, such as the radiation pattern of the loudspeaker itself. The radiation pattern of a loudspeaker is dependent on frequency. The mathematical model of a circular piston model can be seen in the following equation [9]:

$$\mathbf{p}(r,\Theta,t) = \frac{j}{2}\rho_0 c U_0 \frac{a}{r} ka \left[\frac{2J_1(ka\sin\Theta)}{kasin\Theta}\right] e^{j(\omega t - kr)} \tag{4.1}$$

where:

$\mathbf{p}$	=	Pressure	[Pa]
r	=	Distance between the loudspeaker and receiver	[m]
Θ	=	Angle between the loudspeaker orientation and receiver	[rad]
t	=	Time the sound is travelling	[s]
$\rho_0 c$	=	Characteristic Impedance	$[\text{Pa} \cdot \frac{\text{s}}{\text{m}}]$
U_0	=	Particle velocity	$[\frac{\text{m}}{\text{s}}]$
a	=	Radius of the piston	[m]
k	=	Wave number, $\frac{\omega}{c}$	$[\frac{\text{rad}}{\text{m}}]$
ω	=	Angular velocity, $2 \cdot \pi \cdot f$	$[\frac{\text{rad}}{\text{s}}]$
f	=	Frequency	[Hz]
c	=	Speed of sound in the medium	$[\frac{\text{m}}{\text{s}}]$
J_1	=	Bessel function of the first kind of the first order	$[\cdot]$

However, since it is only the radiation pattern that is of interest the model can be simplified to a model very similar to the point source model. This model becomes more directive at the higher frequencies dependent on the size of the loudspeaker. This model can then be formulated as:

$$\mathbf{p}(r,\Theta,t) = \frac{j}{2}\rho_0 c U_0 \frac{a}{r} ka \left[\frac{2J_1(ka\sin\Theta)}{kasin\Theta}\right] \frac{\mathrm{e}^{-j\omega\frac{\|r_b - r_i\|}{c}}}{4\pi\|r_b - r_i\|} \tag{4.2}$$

A simulation of the radiation pattern of a 2 inch loudspeaker can be seen in figure 4.1.

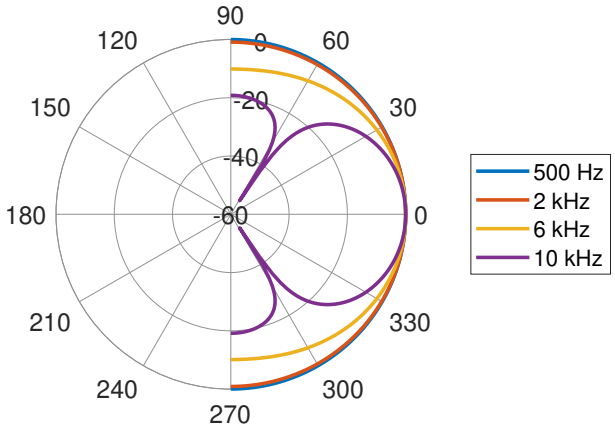


Figure 4.1: Simulation of the radiation pattern for an 2 inch loudspeaker at different frequencies. The model is made with an infinite baffle, thus only 180° are shown. The magnitude is given i normalized dB.

If the wave length is small compared to the size of the loudspeaker, the radiation will be more directive which will decrease the interference between the loudspeakers and can therefore affect the beamforming.

4.2 Physical Placement of the Loudspeakers

The physical placement of the loudspeakers limits the frequencies that can be beamformed. In general, the distance between the loudspeakers, Δ, must be less than half a wavelength at the highest frequency of interest [6]. E.g. beamforming at 1 kHz, the wave length will be 0.343 m thus the spacing between the loudspeakers should at most be 0.1715 m. The distance between the loudspeakers can then determine the limitations of the performance due to spatial aliasing, which is the point where the array can not produce accurate results due to spatial sampling (an analogy could be sampling a digital signal with a too low sampling frequency, aliasing will occur). To avoid spatial aliasing, with a given distance, Δ, between the loudspeakers, the upper frequency limit, f_a, to a given direction θ is given by equation (4.3), where c is the speed of sound in air [31]:

$$f_a \leq \frac{c}{\Delta\left(1 + |\cos(\theta)|\right)} \tag{4.3}$$

From equation (4.3) it can be seen that with a smaller distance between the loudspeakers, a higher frequency limit is achieved before spatial aliasing can be an issue. Also it can be seen that the lowest frequency, with a fixed Δ, is given at 0° and 180°. This is illustrated in figure 4.2, which show f_a as function of the beamforming direction.

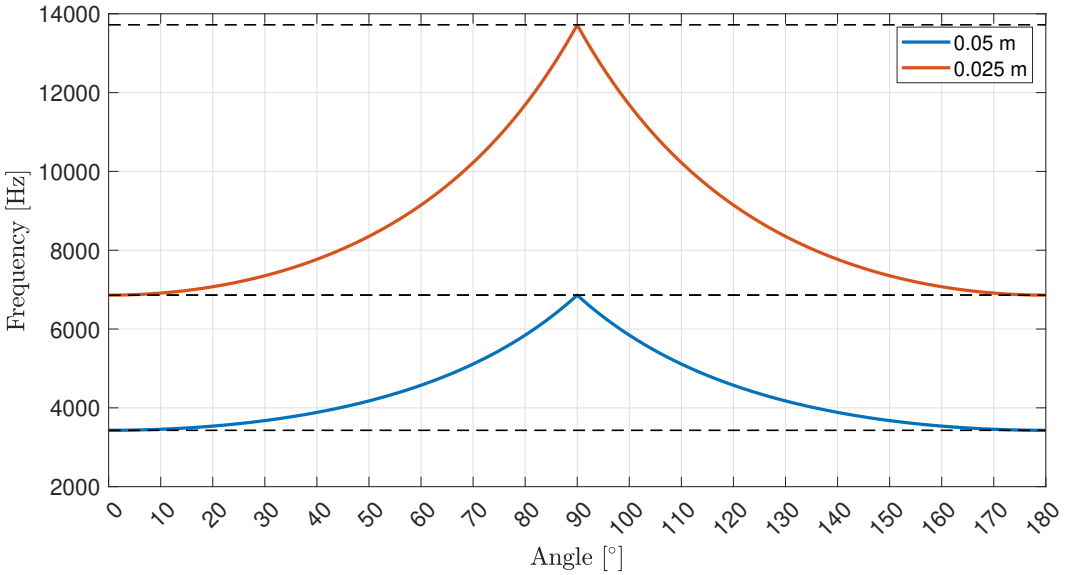


Figure 4.2: Plot of the upper frequency f_a to a given angle before spatial aliasing, shown for $\Delta = 0.05$ m and $\Delta = 0.025$ m. The orientation is that 90° is directly in front of the loudspeaker array.

The lower frequency limit for a loudspeaker array is of course dependent on the loudspeakers frequency range of the loudspeakers, but also the size of the array. The length of the array should be much larger than the wavelength of the lowest frequency used in the beamforming to get a fine angular resolution [6].

4.3 Room Simulation Using the Image Source Method

Since beamforming can be affected by room reflections, a room model, which can be used in the optimization instead of the point source model is described. To simulate the reflections of a room, different methods can be used. The most simple is to make a reverb effect, where a well known implementation of such is the Schroeder reverb [26]. However this implementation will not simulate any reflection but instead just simulate a reverberant effect. Other techniques can be used to simulate reflections of the walls and thereby get the impulse response at a specific point in the room. A simple method is the image source method (ISM) which calculates the reflections for a point source, and thereby the reverberation, in a room for a given sound source and microphone location [32]. The method calculates reflections by mirroring the room at each wall. To illustrate this concept, two reflections to a receiving point r from a sound source radiation from r_s can be seen in a 2D example in figure 4.3. The figure includes a first and a second order reflection.

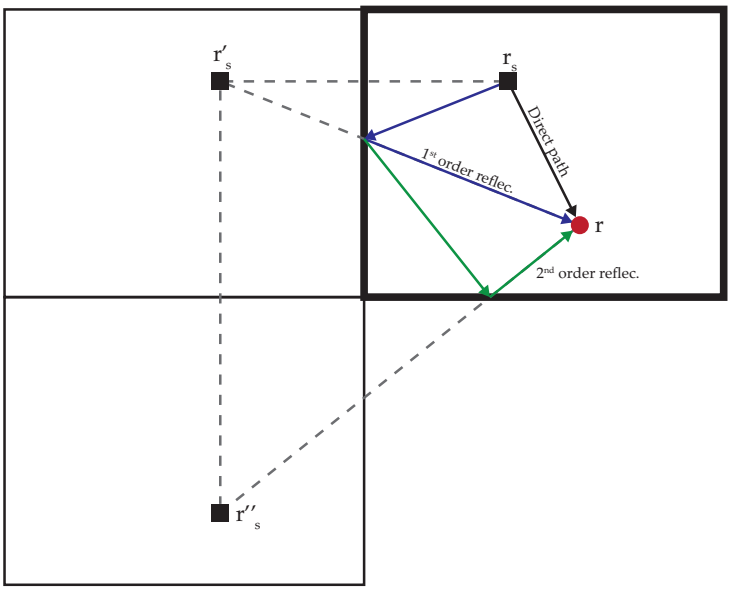


Figure 4.3: Illustration of a first and a second order reflection in the image source method.

To calculate the reflection by the method provided by Allen J. et al. and described in [32], first a room with the dimensions L_x, L_y, L_z is considered, all the measurements are given in meters. The sound source is located in $r_s = [x_s, y_s, z_s]$ and the receiver is located in $r = [x, y, z]$. The coordinate system have its origin, $[0, 0, 0]$, in one of the corners of the room. The locations of the images sources obtained can be expressed as a matrix $\mathbf{R}_p$:

$$\mathbf{R}_p = [x_s - 2p_x x_s - x, y_s - 2p_y y_s - y, z_s - 2p_z z_s - z] \tag{4.4}$$

where $\mathbf{p} = (p_x, p_y, p_z)$ can either take the value 0 or 1. This will result in 8 different combinations, giving a set $\mathcal{P}$, which can be seen illustrated in figure 4.4, where the combinations in each slice through the image space can be seen. The unique combinations are illustrated in green.

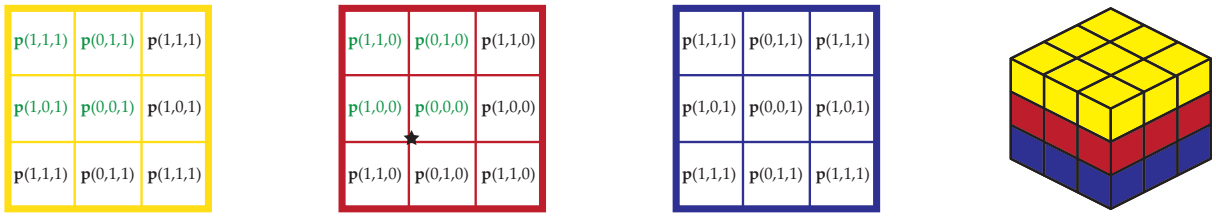


Figure 4.4: Combinations of $\mathbf{p} = (p_x, p_y, p_z)$ for each slice through the image space. The colored cube illustrates the room and the images of it, where each slice corresponds to a layer in the cube. The star indicates the origin.

To take all the reflections into account a matrix $\mathbf{R}_m$ is made, given as:

$$\mathbf{R}_m = 2[m_x L_x, m_y L_y, m_z L_z x] \tag{4.5}$$

where $\mathbf{m} = (m_x, m_y, m_z)$ and each m_x, m_y and m_z can take a value between $-N$ and N, which results in $(2N+1)^3$ combinations which is the number of images sources the algorithm will take into account. Illustration of some of the image sources can be seen in figure 4.5.

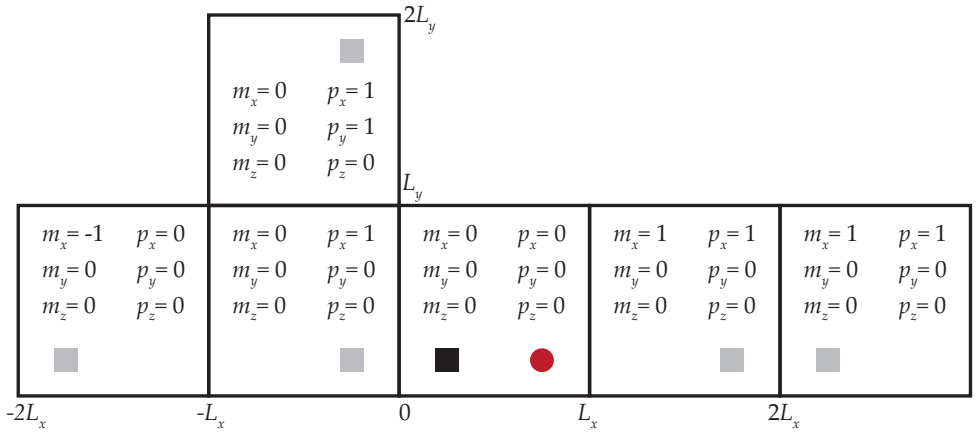


Figure 4.5: Illustration of the structure used to calculate the reflections in the ISM model.

With all the images taken into account the impulse response for the source receiver location can be made by using the distance square law as written in equation (2.4). However, in a room, the different walls can be made of different materials, thus different reflection coefficients should be taken into account. In the proposed method, it is assumed that the reflection coefficient Γ is angle independent. With this assumption the energy absorption coefficient α for a uniform reflection coefficient Γ on a given wall can be described as [32, 33]:

$$\alpha = 1 - \Gamma^2 \tag{4.6}$$

The modification to the impulse response can then be written as in equation (4.7), where c is the speed of sound [32]:

$$h(t, r, r_s) = \sum_{\mathbf{p}\in\mathcal{P}} \sum_{\mathbf{m}\in\mathcal{M}} \Gamma_{x1}^{|m_x - p_x|} \Gamma_{x2}^{|m_x|} \Gamma_{y1}^{|m_y - p_y|} \Gamma_{y2}^{|m_y|} \Gamma_{z1}^{|m_z - p_z|} \Gamma_{z2}^{|m_z|} \cdot \frac{\delta(t - (\|\mathbf{R}_m + \mathbf{R}_p\| / c))}{4\pi \|\mathbf{R}_m + \mathbf{R}_p\|} \tag{4.7}$$

The different Γ's are the pressure reflection coefficients for the six boundaries in the room, where the subscript indicates the wall. It should be noted that these coefficient normally are frequency dependent, which is not a part of this algorithm.

The MATLAB implementation of this algorithm can be found in appendix B. A simulation of a room have been made with two different values of α. The results can be seen in figure 4.6 and 4.7, where the calculated impulse can be seen next to the decay curve. The decay curve is made with Schroeder's backward integration, and the reverberation time can be found using T60, both described in [34, 35].

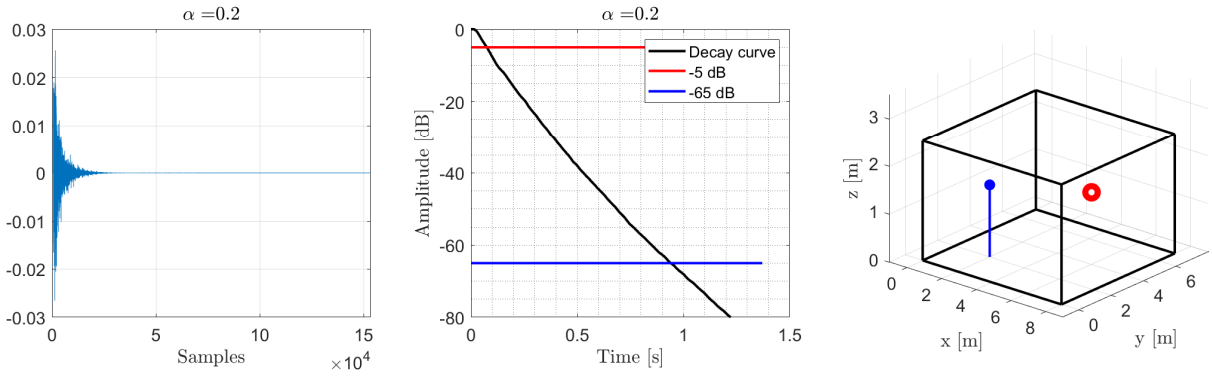


Figure 4.6: Impulse response and decay curve for a room, calculated using ISM. The parameters are as followed: $[L_x, L_y, L_z] = [8, 7, 2.5]$, $r = [2, 2, 1.5]$, $r_s = [6, 4, 1.5]$, $f_s = 48$ kHz and $N = 100$. In the plot to the right, the room is plotted, where the blue dot are the receiver and the red circle are the source.

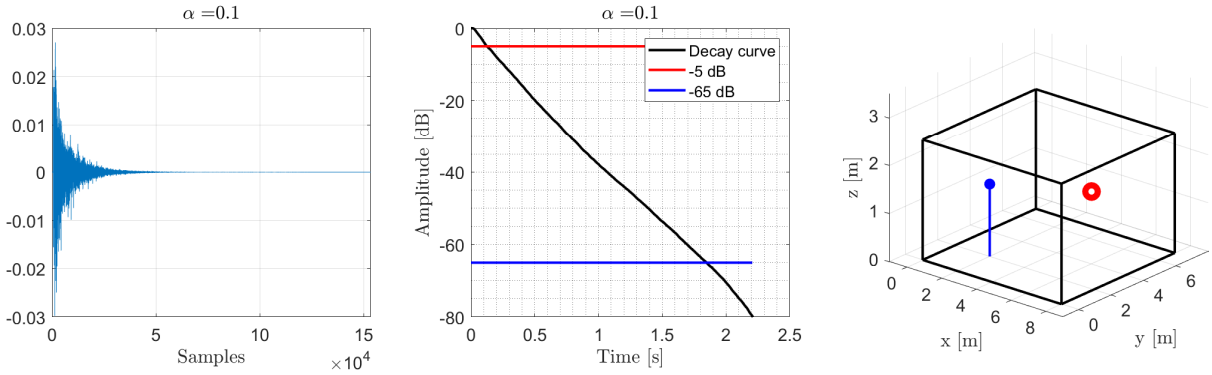


Figure 4.7: Impulse response and decay curve for a room, calculated using ISM. The parameters are as followed: $[L_x, L_y, L_z] = [8, 7, 2.5]$, $r = [2, 2, 1.5]$, $r_s = [6, 4, 1.5]$, $f_s = 48$ kHz and $N = 100$. In the plot to the right, the room is plotted, where the blue dot are the receiver and the red circle are the source.

The mean reverberation time is 0.87 s for $\alpha = 0.2$ (figure 4.6) and 1.71 for $\alpha = 0.1$ (figure 4.7).

4.4 Measurement Setup

The measurement setup used, contains a loudspeaker array with a total of 39 loudspeakers, each of them spaced 10 cm apart horizontally, two arrays are placed on top of each with a horizontal offset of 5 cm, giving an effective spacing in the horizontal plane of 5 cm between each loudspeaker. A sketch of the setup is seen in figure 4.8, and panorama pictures of the room can be seen in figures 4.9 and 4.10.

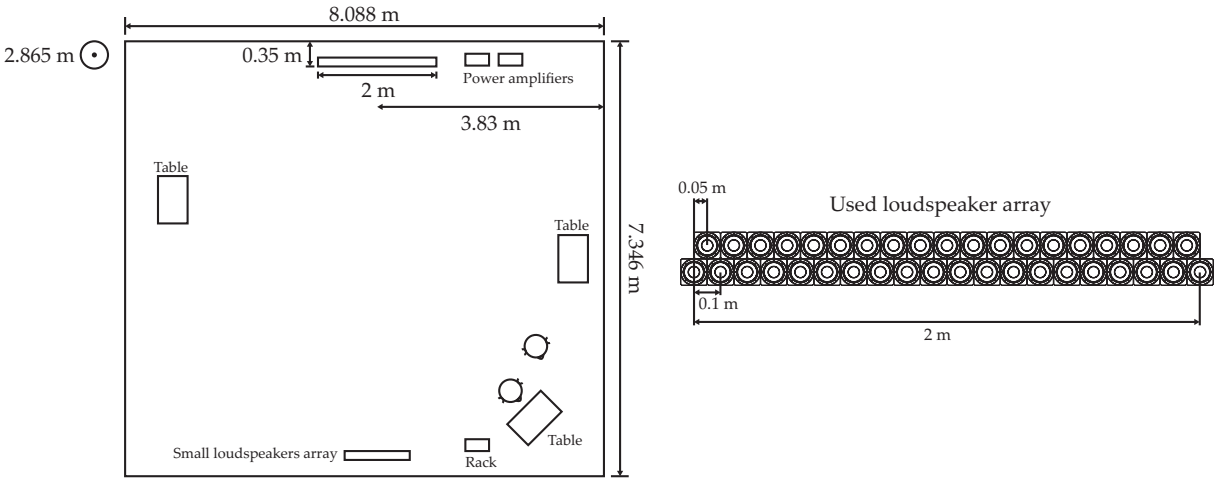


Figure 4.8: Sketch of the room.

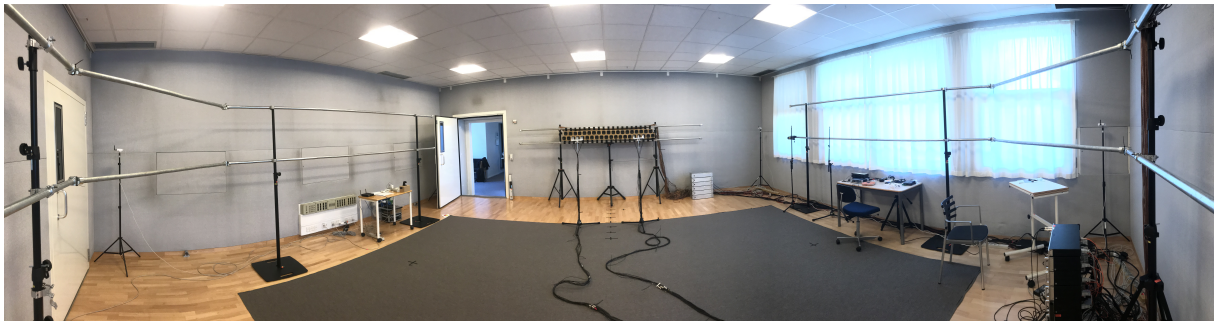


Figure 4.9: Panorama picture of the room.

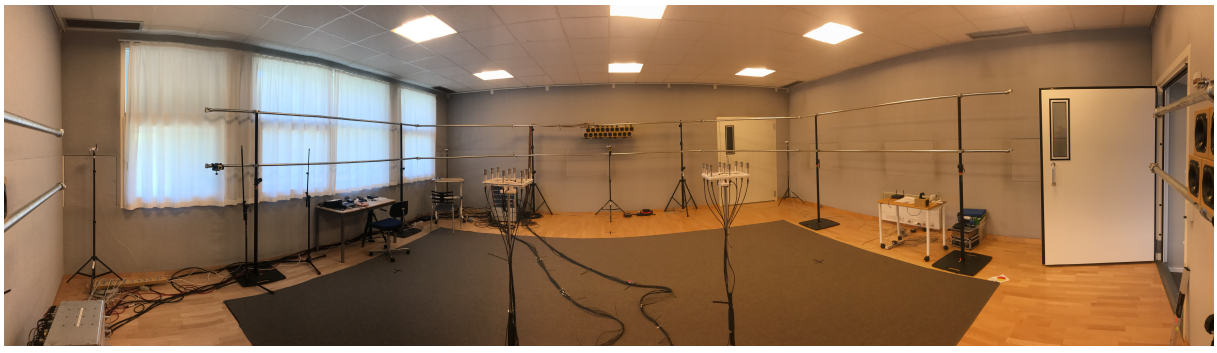


Figure 4.10: Panorama picture of the room.

For the optimization 10 control points have been used, 1 in the bight zone and 9 in the dark zone. However, in order to evaluate how the sound recreation is surrounding the bright control point, a total of 18 measurement points were used. A grid mount have been made in order to ensure that the microphones are place correctly relative to each other, and to enable easy adjustments of the bright and dark zone center position. The mount is shown in figure 4.11.

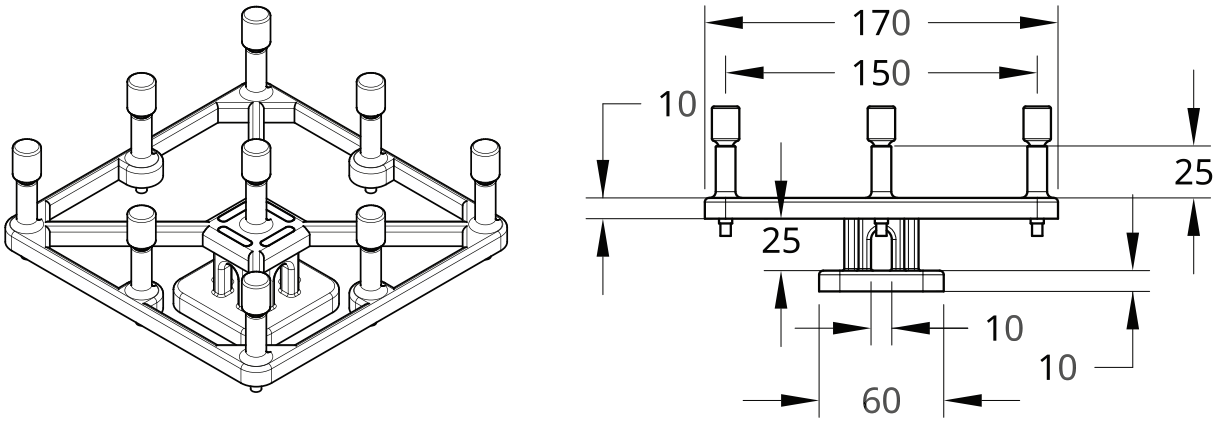


Figure 4.11: Model of 3D-printed microphone mount. Dimensions are given i [mm].

The hardware setup were built up by the use of: 18× GRAS 40AZ, 39× midrange loudspeaker units, Fireface UFX+, 2× RME Micstasy, 3× RME M-16 DA and 6× 150 W ICEPOWER amplifiers. They were all connected through the Multichannel Audio Digital Interface (MADI). The setup is illustrated in figure 4.12. A full measurement journal can be found in appendix C.

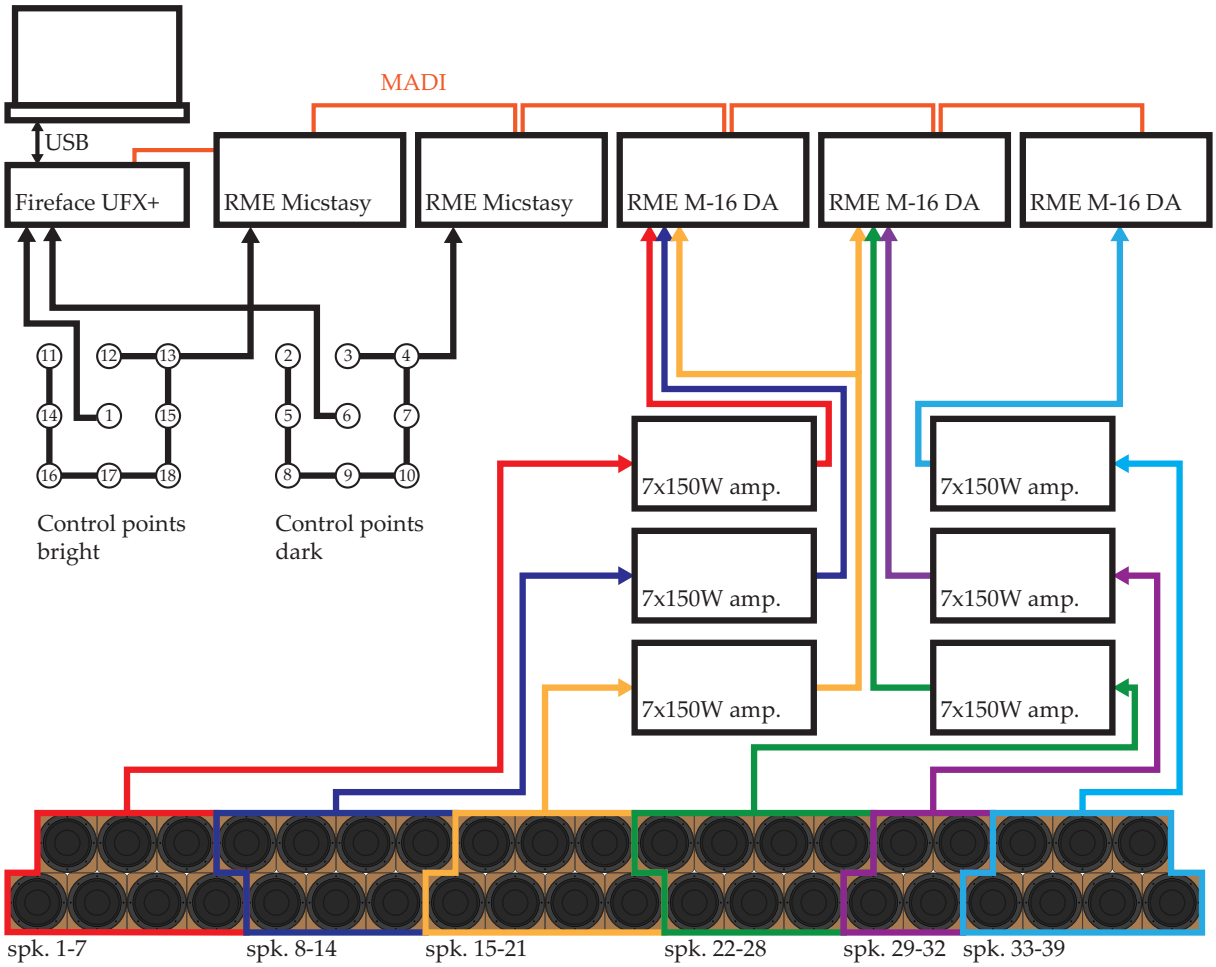


Figure 4.12: Hardware connection diagram of the system used for measurements.

5 | Test and Validation of Simulation Model

In this chapter, a validation of the simulation model used so far will be made. This is done by using measured impulses responses from a real test setup in a real room. After validating the simulation model, different settings will be simulated before a final test. These simulations will include the two additional models (piston model and ISM) of the acoustic transfer function (ATF), used in place of the point source model. Afterwards a comparison will be made between the simulated and the measured results. Finally some additional simulations testing different values of λ based on the results of the measurements will be made.

5.1 Measuring Impulse Responses Used for Simulations

First some measurements are made in order to test how the estimated ATF-models compare with the measured one. A close realisation of the actual room used for measurement can be made, by measuring the room impulse responses (RIR) for the room at the position of the control points. A convolution of the RIR's with a desired signal can be used as a baseline of how the system is to expected behave in actual measurements. Afterwards, both the simulated and measured contrast and MSE will be looked at. This is done for two different setups of the bright zone and dark zone position.

The first setup is with the worst assumed position where the two sound zones are placed directly in front of each other with the dark zone blocking the direct path between the bright zone and the loudspeaker array, and with every loudspeaker being closer to the dark zone control points than the bright zone. The bright zone center is placed at $[0, 1.5]$ and dark zone center is placed at $[0, 0.5]$. This will be referred to as **setup 1**. The other setup is similar, however in this setup the zones have been moved away from the middle, where the bright zone center is placed at $[0.25, 1.5]$ and dark zone center is placed at $[-0.25, 0.5]$. This will be referred to as **setup 2**. The loudspeaker array and all of the control points are placed in a height of 1.485 m from the floor. The two setups can be seen illustrated in figure 5.1.

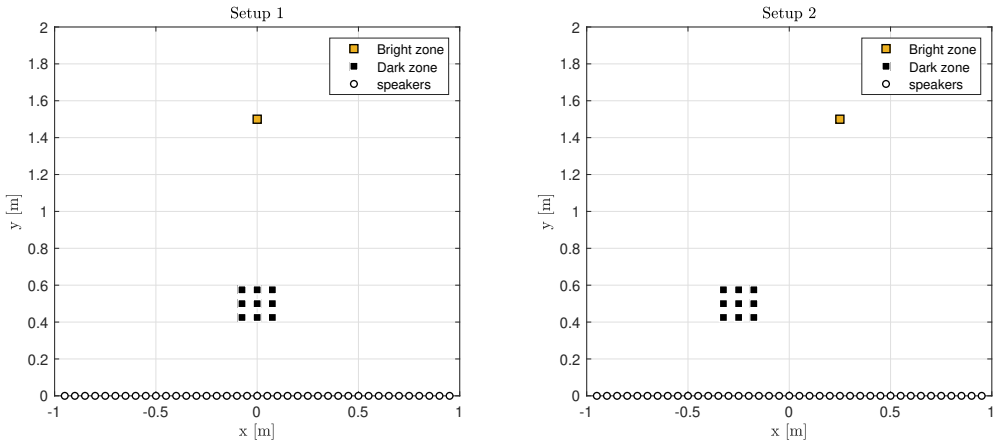


Figure 5.1: Plot of the control points positions relative to the loudspeaker array.

To create the impulse responses for all the control point loudspeaker combinations the Swept Sine Method have been used [36]. A description of the method and how it is used can be found in appendix D. When using the Swept Sine Method, it can be advantageously to utilize its ability to suppress uncorrelated noise, as a big part of the the additional noise on the recordings are assumed to be uncorrelated. This can be done by taking the average of several impulse measurements resulting in a higher signal-to-noise ratio (SNR) [37], for these measurements the number of recordings for each combination of control points and loudspeakers were chosen to be five. A simulating of this can be seen in appendix D figure D.2. The input signal is a logarithmic chirp from 20 Hz to 24 kHz (half of the sampling frequency 48 kHz). The chirp is then windowed with a shifted Hanning window with a length of 1.5 seconds. This is done in order to avoid too much energy at the lower frequencies. The duration of the signal is 7 seconds with 1 second pre-delay, 5 seconds chirp and 1 second post-delay.

The recorded signals are filtered with a second order high pass Butterworth filter with cutoff frequency at 50 Hz, before making the impulse responses. This is due to a continuous low frequency noise introduced from the microphones, this can be seen in appendix D.

An example of one of the created impulse responses can be seen in figure 5.2.

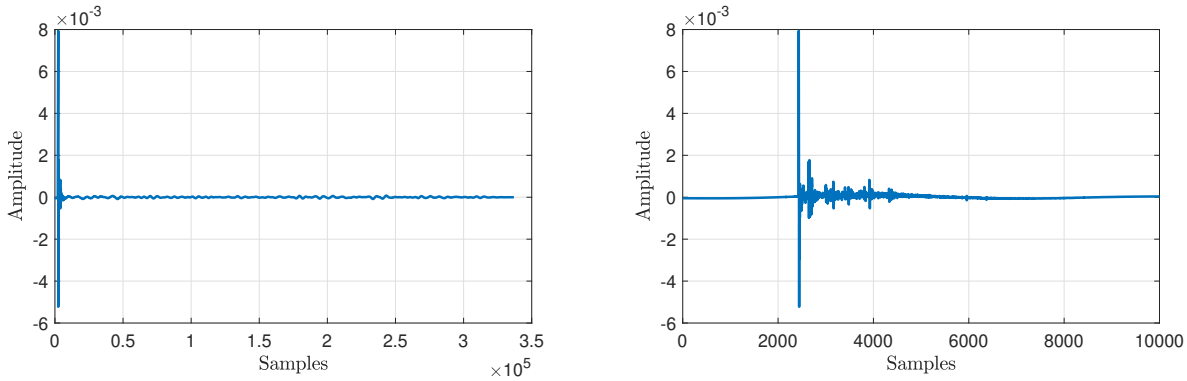


Figure 5.2: Example of an impulse response created using the Swept Sine Method.

Looking at the impulse it can be seen that the pulse spikes after around 2000 samples. With a

sampling frequency of 48 kHz this delay will corresponds to a distance, between the loudspeaker and the microphone, of approximately 14 m, which is not the case. The additional delay in the impulse response is caused by the system delay. To find the system delay, a simple test have been made. Connecting the output of the DA-converter directly to the input of the AD-converter where a pulse was then played while recording. The results can be seen in figure 5.3.

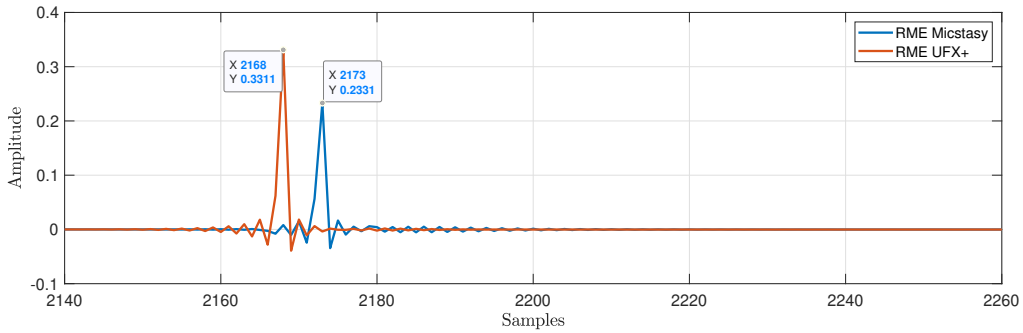


Figure 5.3: Results for test of system delay while simultaneously recording and playback.

As it can be seen in figure 5.3, the system delays are 2168 for the RME UFX+ and 2173 samples for the RME Micstasy, which then can be removed from the created impulse responses.

5.2 Comparison of Different Acoustic Transfer Functions Used for the Simulation Model

A validation of the simulation model have been performed in order to see how precise results that can be expected. The validation is done as a comparison between the simulated frequency response and the measured frequency response. Both using the same input signal. In order to be able to compare the signals both the simulation model and the measured data have been calibrated to 94 dB ref. 20 μPa. Additionally a plot of the played signal convoluted with the measured room impulse responses have been made. The amplitude response is expected to be really close to the measured signal. The following results are shown in 1/3 octave bands plots. This is primarily done because of readability. Additionally it show a better representation of how sound is perceived [24].

The ISM model have been tweaked so that the reverberation time match the measured reverberation time of the measurement room, found as the mean of all the loudspeaker/control point combinations and the mean of the frequencies from 200 - 24000 Hz. The reverberation time is found to be 0.18 s, where the procedure is described in appendix D.

Figures for each of the simulation models are shown both for the bright zone control point on the left as well as the dark zone center control point on the right. For these tests the played signal was a sweep from 300 - 24000 Hz, with a pre-delay of 1 second and a post delay of 2 seconds. The figures show the sound pressure level (SPL) from 315 - 4000 Hz, based on the upper frequency limit for spatial aliasing, and the lower frequency limit based on the size of the loudspeaker array.

Setup 1

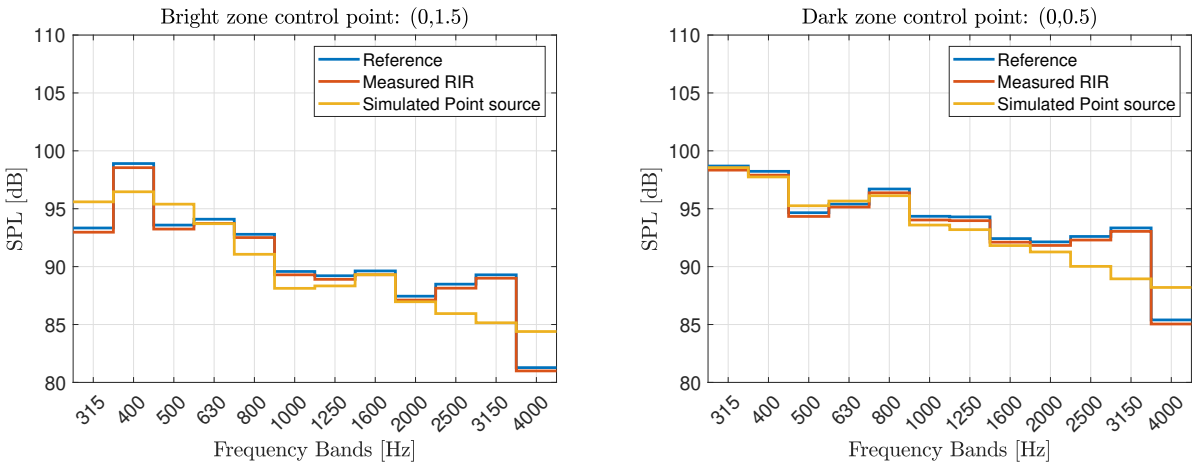


Figure 5.4: Comparison of the reference measurement of the test-sweep, the test-sweep convoluted with the measured impulse response and the test-sweep convoluted with the simulated point source free field impulse response, for **setup 1**.

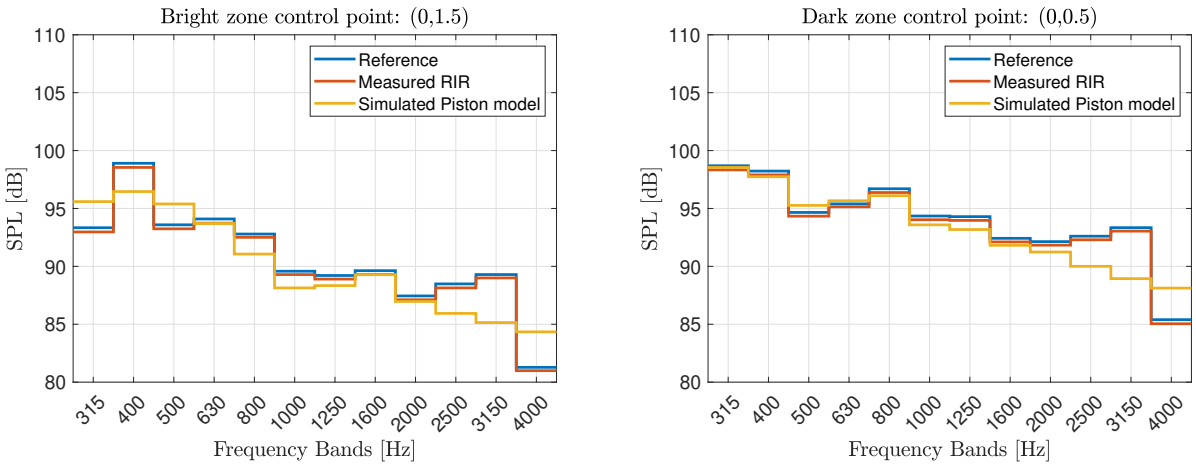


Figure 5.5: Comparison of the reference measurement of the test-sweep the test-sweep convoluted with the measured impulse response and the test-sweep convoluted with the simulated piston model free field impulse response, for **setup 1**.

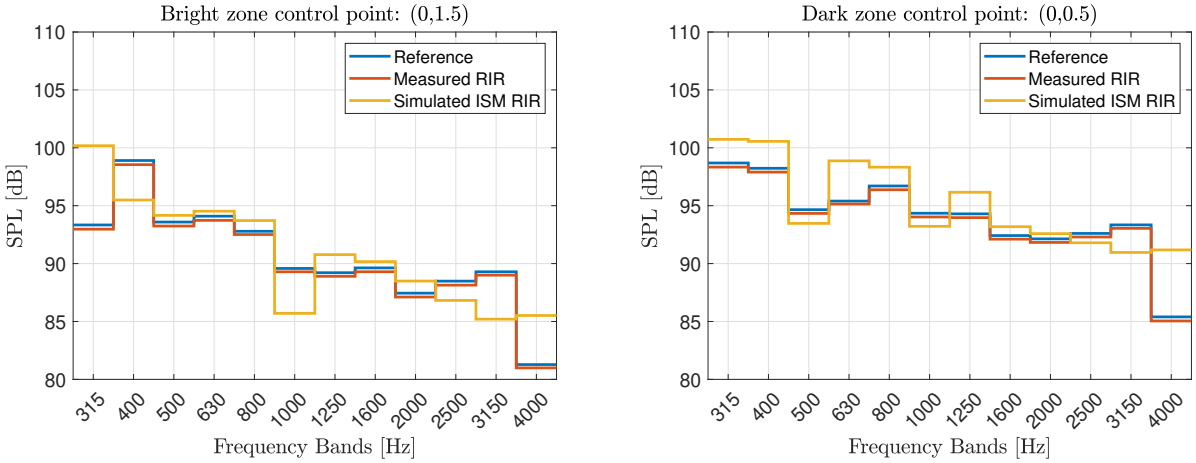


Figure 5.6: Comparison of the reference measurement of the test-sweep the test-sweep convoluted with the measured impulse response and the test-sweep convoluted with the simulated ISM-room impulse response, for **setup 1**.

Setup 2

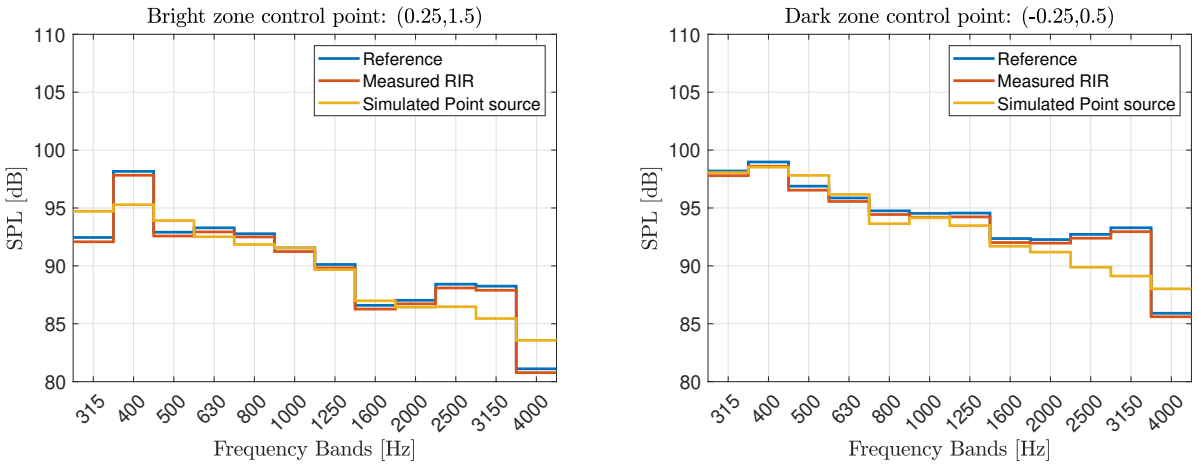


Figure 5.7: Comparison of the reference measurement of the test-sweep the test-sweep convoluted with the measured impulse response and the test-sweep convoluted with the simulated point source free field impulse response, for **setup 2**.

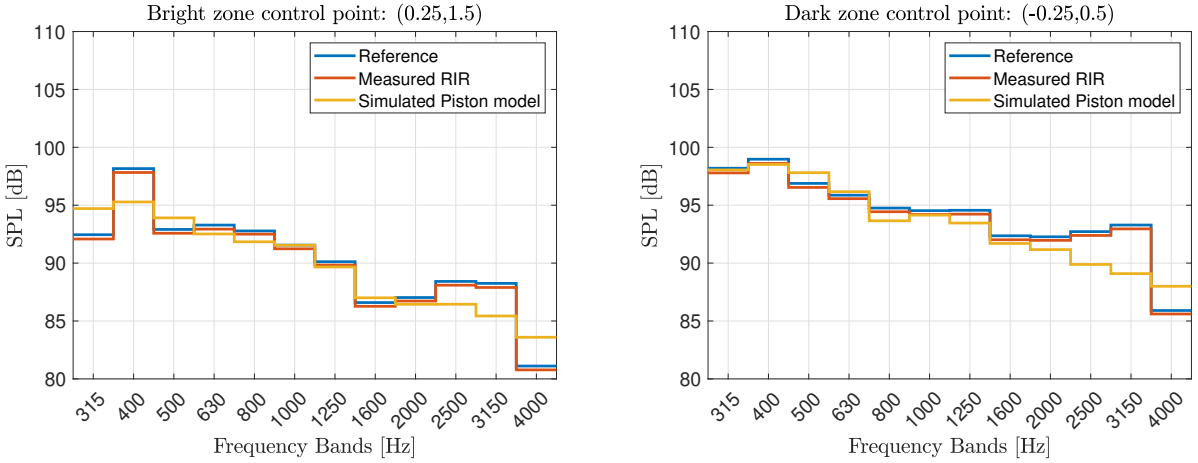


Figure 5.8: Comparison of the reference measurement of the test-sweep the test-sweep convoluted with the measured impulse response and the test-sweep convoluted with the simulated piston model free field impulse response, for **setup 2**.

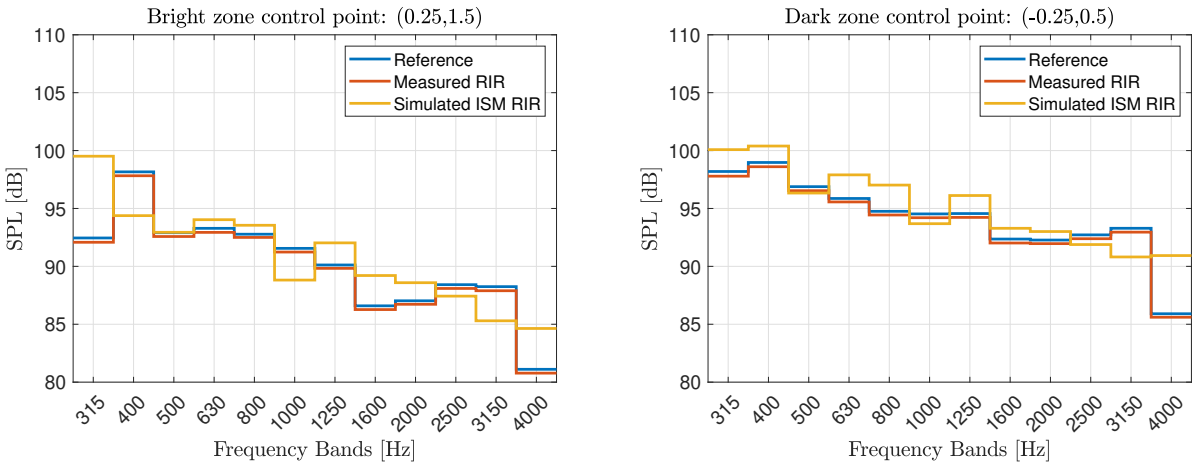


Figure 5.9: Comparison of the reference measurement of the test-sweep the test-sweep convoluted with the measured impulse response and the test-sweep convoluted with the simulated ISM-room impulse response, for **setup 2**.

Both the point source model and the piston model show an almost identical performance, both of them showing a close match at the frequencies between 500 and 2000 Hz, where they differentiate from the measured signal a bit more at the other frequencies, but does however still come close to the measured signal. In general it can be seen that the simulated ISM RIR model, is still somewhat close to the measured signal, but does fluctuate a bit more. This is probably due to the simplicity of the reflections in this model where the reflection coefficients are not frequency dependent, which they are expected to be in the real measurement. Because of how attenuated the measurement room is, this could explain the good performance of the free field models. It can also be seen that in every figure, a very close match is achieved between the measured signal and the signal that have been convoluted with the measured impulse response. Because of this, the impulse responses convoluted with the filtered signal achieved from the simulation algorithm can be used to investigate the results of the simulation of the contrast measurements before the actual measurements.

5.3 Comparison of the Achieved Contrast Using the Different Simulation Models

To compare the different ATF models performance, simulations have been made using the measured impulse responses for the room. This is done in order to get as close to the real scenario as possible. The simulations have been made for the two different setups described earlier. **Setup 1** where the bright control point is placed in $[0, 1.5]$ and the dark control points are placed with center in $[0, 0.5]$. **Setup 2** is where the bright control point is placed in $[0.25, 1.5]$ and the dark control points are placed with center in $[-0.25, 0.5]$. These positions are both relative to the center of the loudspeaker array.

The filters are calculated for the different ATF models using the wPM algorithm. The initial settings of the algorithm are: $\xi = 0.9$, $\lambda = 0.1$ and the filters are calculated for 1 Hz to 24 kHz in steps of 1 Hz. The calculated filters are made into impulse responses which are then convoluted with a test signal, this test signal is a logarithmic chirp from 300 Hz to 24 kHz. The filtered test signals are then convoluted with the measured impulse responses in order to simulate the pressure in a given point. The contrasts are found be subtracting the pressure in the bright zone control point with the mean pressure in the dark zone control points.

To recap, the different ATF models are: *point source*, *piston model*, *Estimated RIR* (output from the ISM model) and *Measured RIR*. The last two models are given in the time domain as impulse responses, thus an FFT is used so that the wPM algorithm can use them as an input. The different ATF models, can be seen in figure 5.10. For readability, the inputs (the transfer functions) are shown as amplitude responses. These responses are shown for a single loudspeaker control point combination. The control point is placed in $[0, 1.5]$ and the loudspeaker is placed with center in $[-0.95, 0]$.

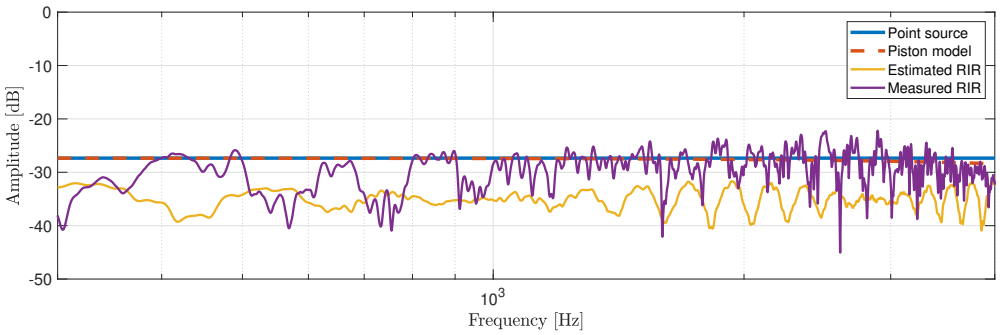


Figure 5.10: ATF input to the algorithm for one source control point combination showed as an amplitude response in the frequency domain showed for relevant frequencies.

As it can be seen in figure 5.10, the amplitudes of the different ATF models are somewhat in the same region but both the estimated and measured RIR are more fluctuating, compared to that of the point source and piston model.

In the following two subsections, the different models will be tested in the two different setups. This will be done using different settings of the ray space edge subdivision (RSES) model, as well as a reference where standard wPM is used. Here both the MSE and the contrast will be

used to compare the performance of the difference settings. To calculate the MSE a test signal is made as a single impulse which will be the desired signal. The filters are convoluted with a test signal, and then with the impulse responses of the room. The MSE is then calculated in the frequency domain between the desired signal and the simulated signal. The contrast is calculated as the mean pressure difference between the bright zone and the dark zone, for this, a logarithmic chirp from 300 - 24000 Hz have been used.

Setup 1

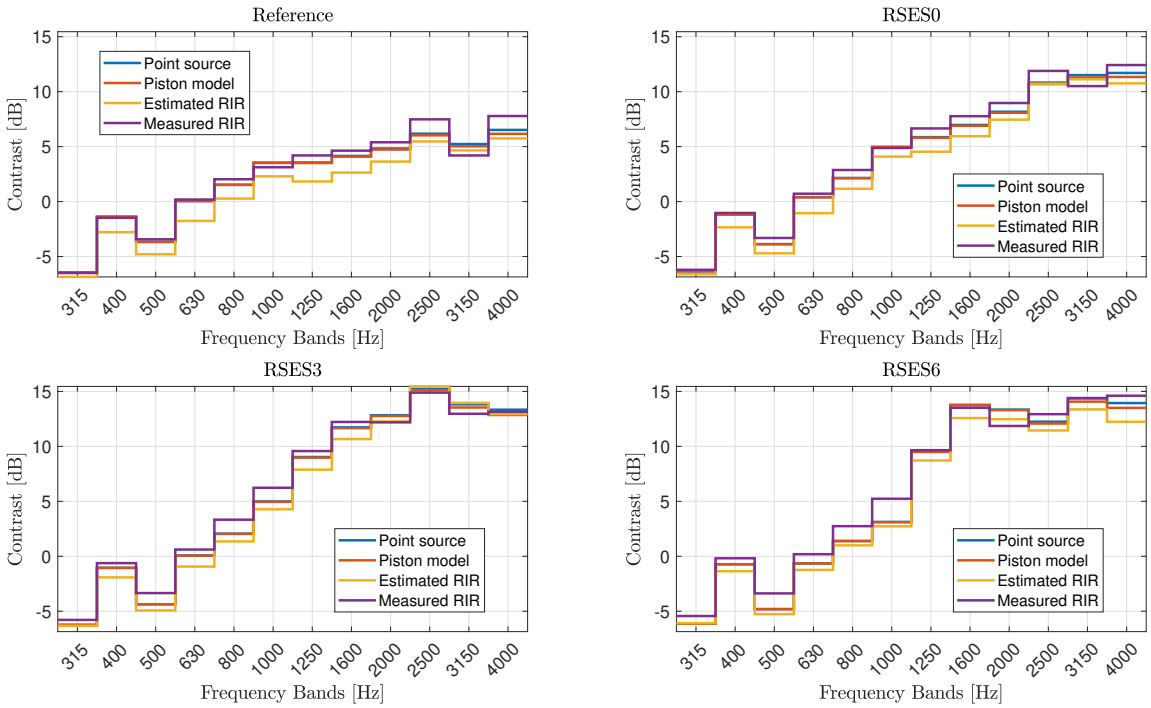


Figure 5.11: Comparison of different models performance in order to achieved contrast when convolving the calculated filters with the measured impulse responses in the room. The setup is the one described in section 5.1, where the placement of the microphones are, relative to the loudspeaker array, $[0, 1.5]$ for the bright zone center and $[0, 0.5]$ for the dark zone center.

It can be seen in figure 5.11 that the different models perform close to the same when looking at the achieved contrast. The point source model and the piston model show almost identical performance, which make sense since the piston model of a 2 inch loudspeaker is very close to have an omnidirectional polar pattern in the shown frequency area. This can also be seen in figure 5.10, where only a small difference is seem at the highest frequencies. In this setup, the model that performed the worst is the estimated RIR.

For readability the plots of the different ATF-models with different settings of the RSES-model have been merged into one plot. This is done in order to see how the RSES-model affects the different ATF-models. The results can be seen in figure 5.12.

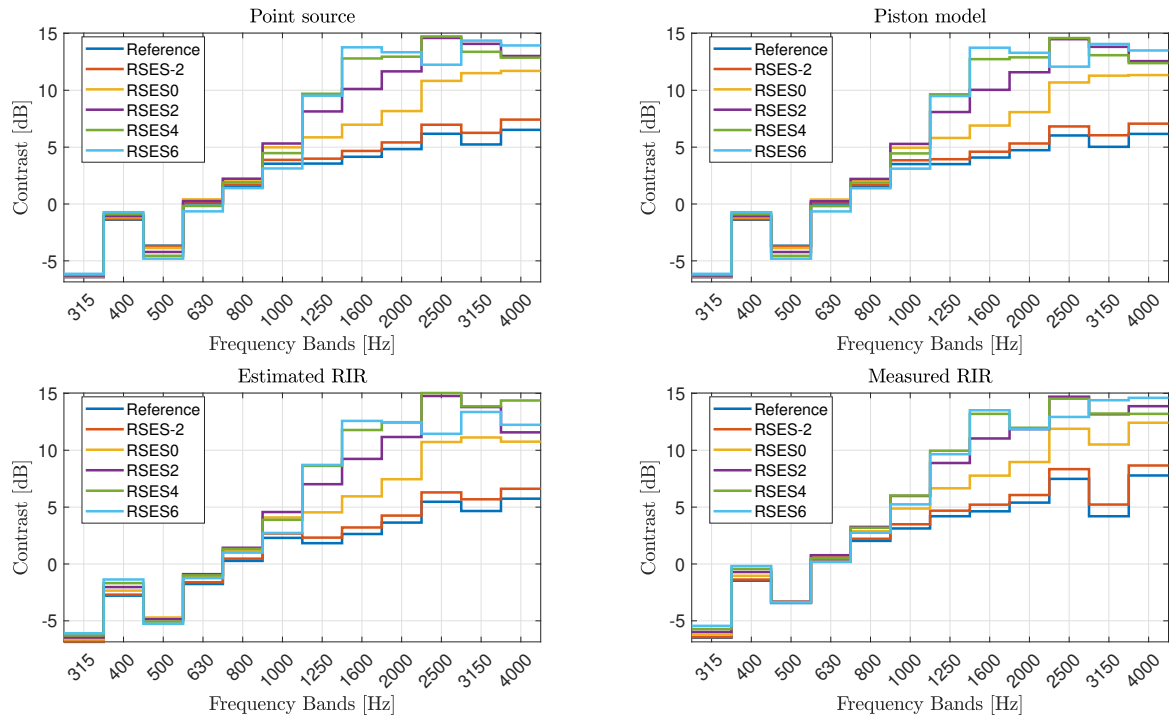


Figure 5.12: Showing the effect of using the RSES-model with different settings to turn off loudspeakers. The placement of the microphones are, relative to the loudspeaker array, $[0, 1.5]$ for the bright zone center and $[0, 0.5]$ for the dark zone center.

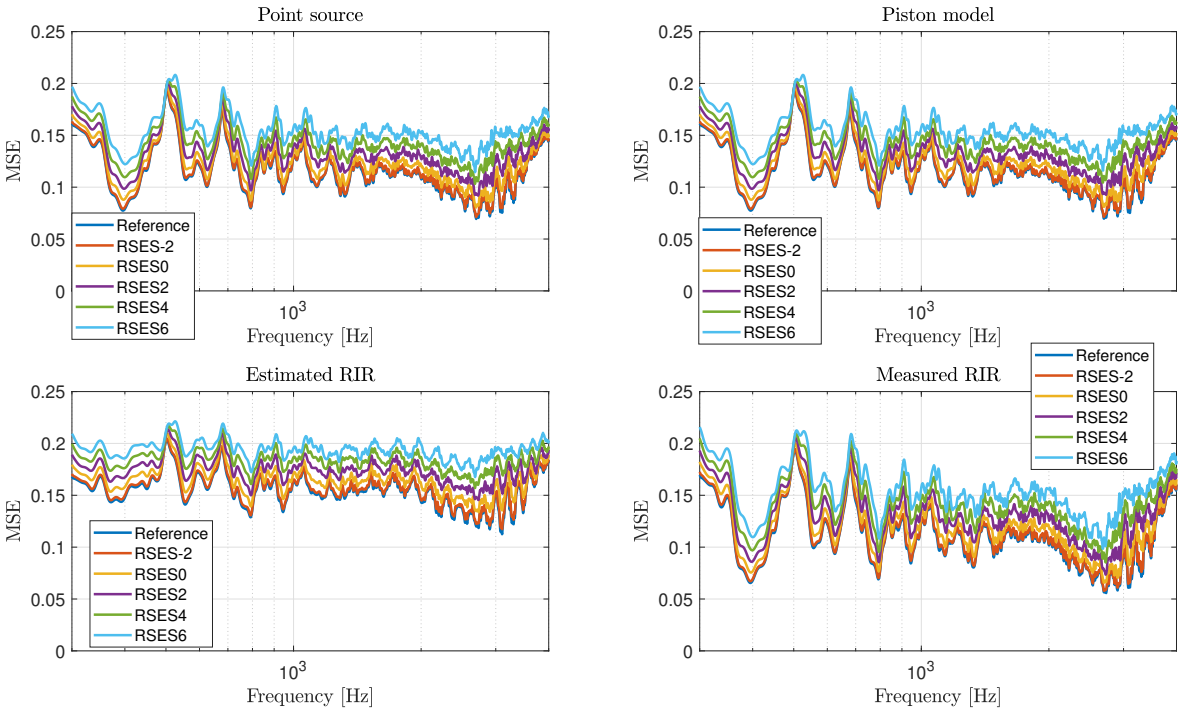


Figure 5.13: Showing the effect on the MSE of using the RSES-model with different settings to turn off loudspeakers. The placement of the microphones are, relative to the loudspeaker array, $[0, 1.5]$ for the bright zone center and $[0, 0.5]$ for the dark zone center.

It can be seen in figure 5.12 that using the RSES-model to remove loudspeakers, can achieve a higher contrast. It seems that turning off additional loudspeakers does further increase this contrast, but seems less effective the more loudspeakers that are being turned off. Looking at the MSE it can be seen that the more loudspeakers removed the higher the MSE, shown in figure 5.13. Therefore the tradeoff for achieving higher contrast is a higher MSE, which was also shown earlier in sections 3.3 and 3.5.

Setup 2

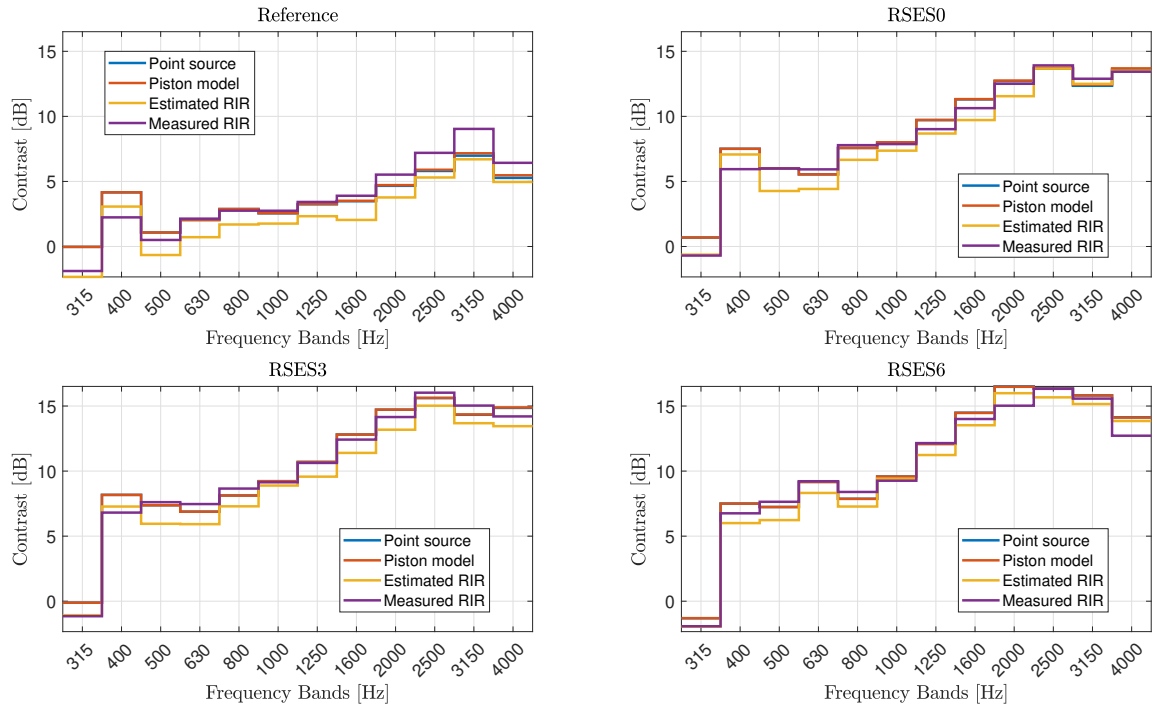


Figure 5.14: Comparison of different models performance in order to achieved contrast when convolving the calculated filters with the measured impulse responses in the room. The setup is the one described in section 5.1, where the placement of the microphones are, relative to the loudspeaker array, $[0.25, 1.5]$ for the bright zone center and $[-0.25, 0.5]$ for the dark zone center.

Again in this setup, it can be seen that the models are very close to each other, and the contrast increases when using the RSES-models. Once again the plots of the different ATF-models with different settings of the ray space subdivision model have been merged into one plot in order to see how the RSES-model affects the different ATF-models. The results can be seen in figure 5.12 and the MSE can be seen in figure 5.16..

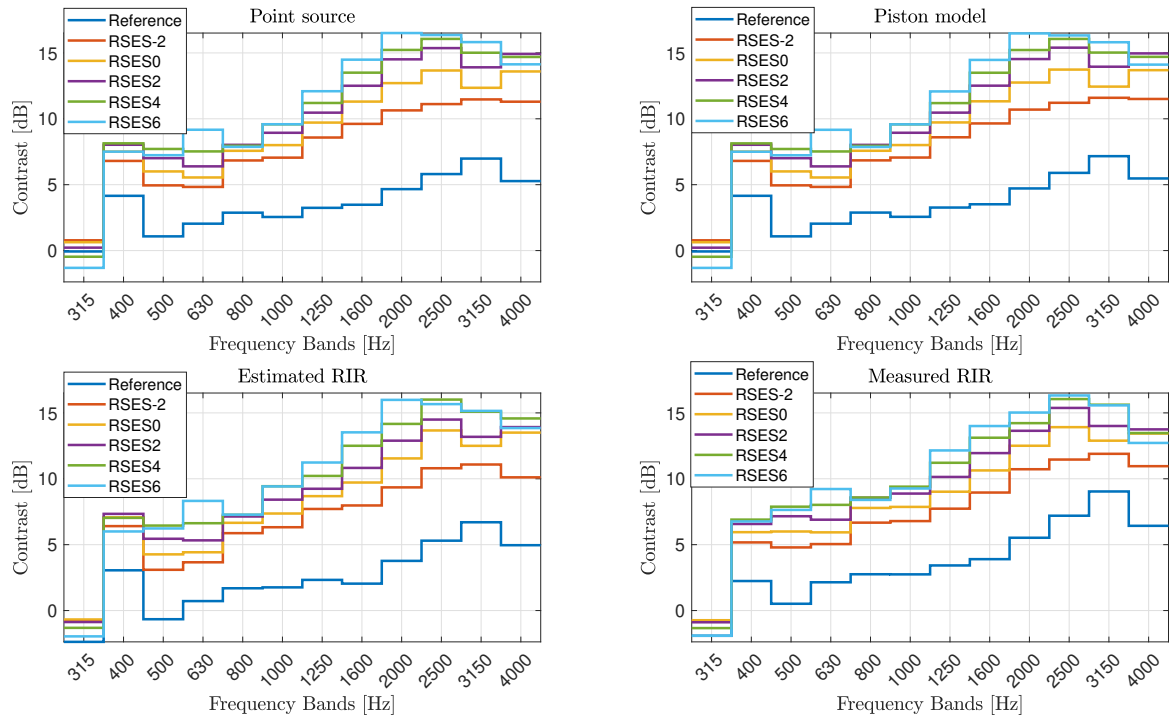


Figure 5.15: Showing the effect of using the RSES-model with different settings to turn off loudspeakers. The placement of the microphones are, relative to the loudspeaker array, $[0.25, 1.5]$ for the bright zone center and $[-0.25, 0.5]$ for the dark zone center.

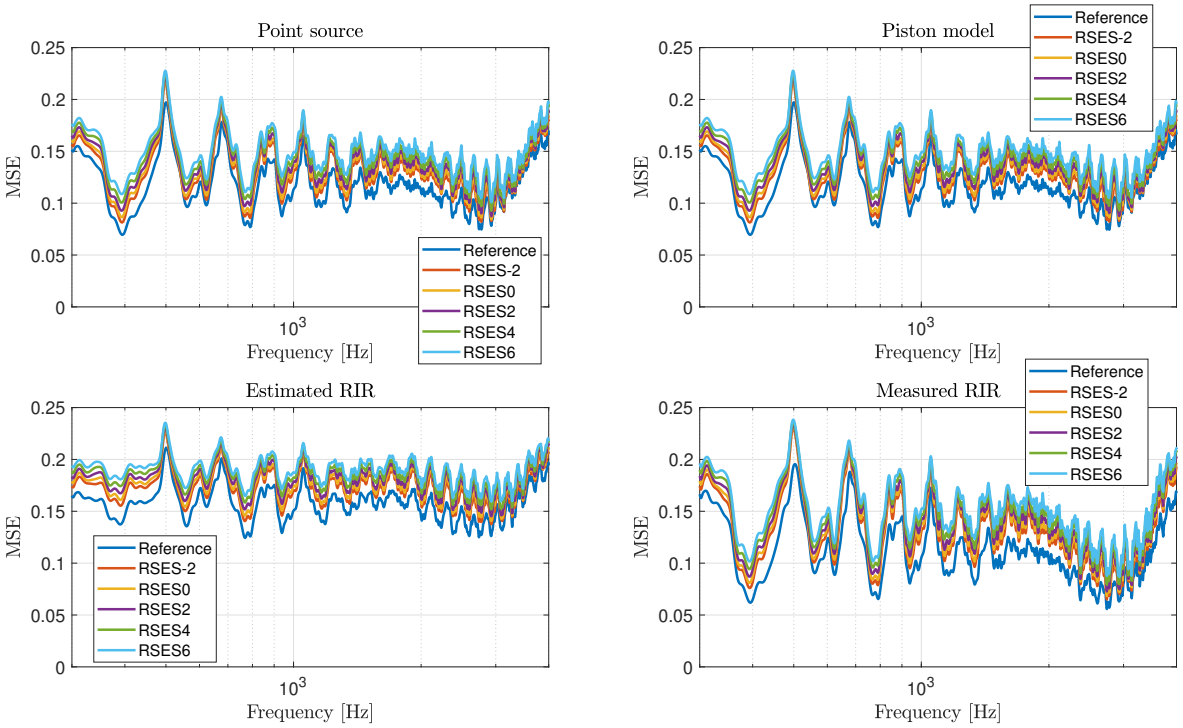


Figure 5.16: Showing the effect on the MSE of using the RSES-model with different settings to turn off loudspeakers. The placement of the microphones are, relative to the loudspeaker array, $[0.25, 1.5]$ for the bright zone center and $[-0.25, 0.5]$ for the dark zone center.

In figure 5.15 it can be seen that **setup 2** follows the same tendency as **setup 1**. When using the RSES-model to turn off loudspeakers does increase the contrast and that the additional contrast achieved from removing additional loudspeakers is getting less effective if more loudspeakers gets turned off. The MSE can be seen in figure 5.16. Where it again show that less loudspeakers does result in a higher MSE. In figure 5.17, the mean MSE for the whole signal can be seen for the different ATF-models and difference RSES settings, for an easier comparison.

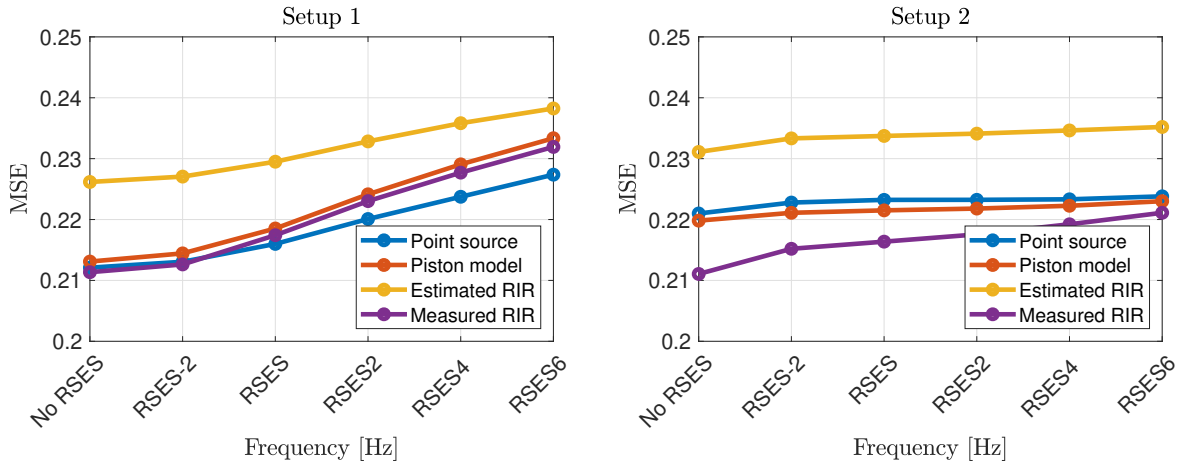


Figure 5.17: Mean MSE for the entire signal with different ATF-models and different RSES settings.

This does in general show best performance from the optimization made with the measured RIR and the point source and piston model. Where the estimated RIR for both cases is the worst model. Due to how similar the piston model and the point source model are in terms of performance, the point source model will be used for further simulation due to the simplicity of this model.

5.4 Regularization Parameter λ's Impact on the Ray Space Subdivision Model

To test how the regularization parameter λ affect the ray space subdivision model, simulations have been made. λ control the array effort which describes how much energy that can be supplied to the loudspeakers. The tests are made with four different values of λ, simulated for **setup 1** and **setup 2** respectively. The results are only showed with the optimization done with the point source model. This is chosen due to the very similar results from the other models when decreasing the value of λ.

Setup 1

Different values of λ have been tested for **setup 1**, in figure 5.18. Different RSES settings can be seen with different settings of λ. In figures 5.19 and 5.20, the filters in frequency can be seen for the different values of λ. Here the reference and the RSES6 setting is shown.

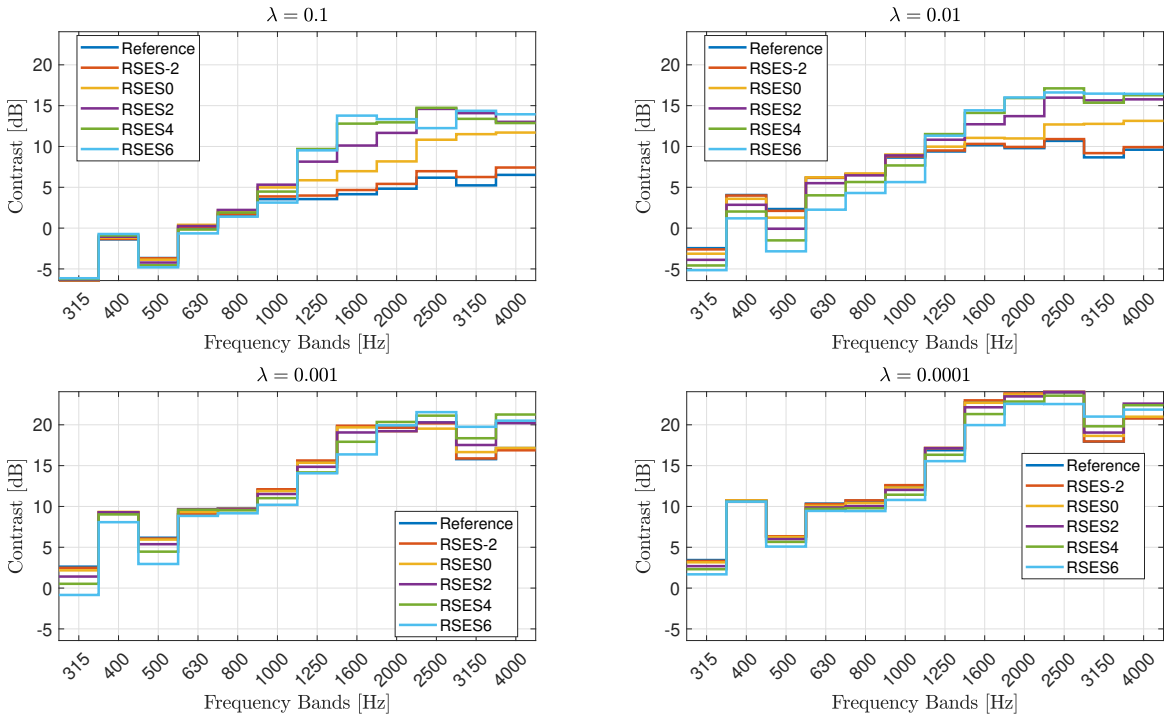


Figure 5.18: Contrast between the bright and dark control points, and the effect of using different values of λ in **setup 1**.

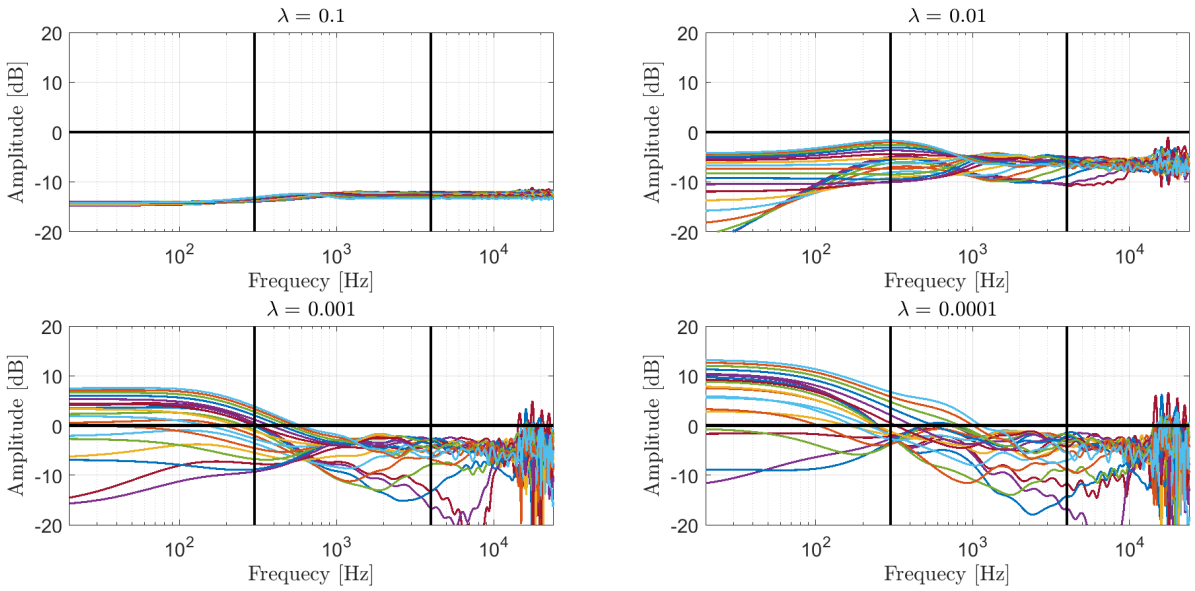


Figure 5.19: The amplitude response of the reference filters for **setup 1**, with different values of λ.

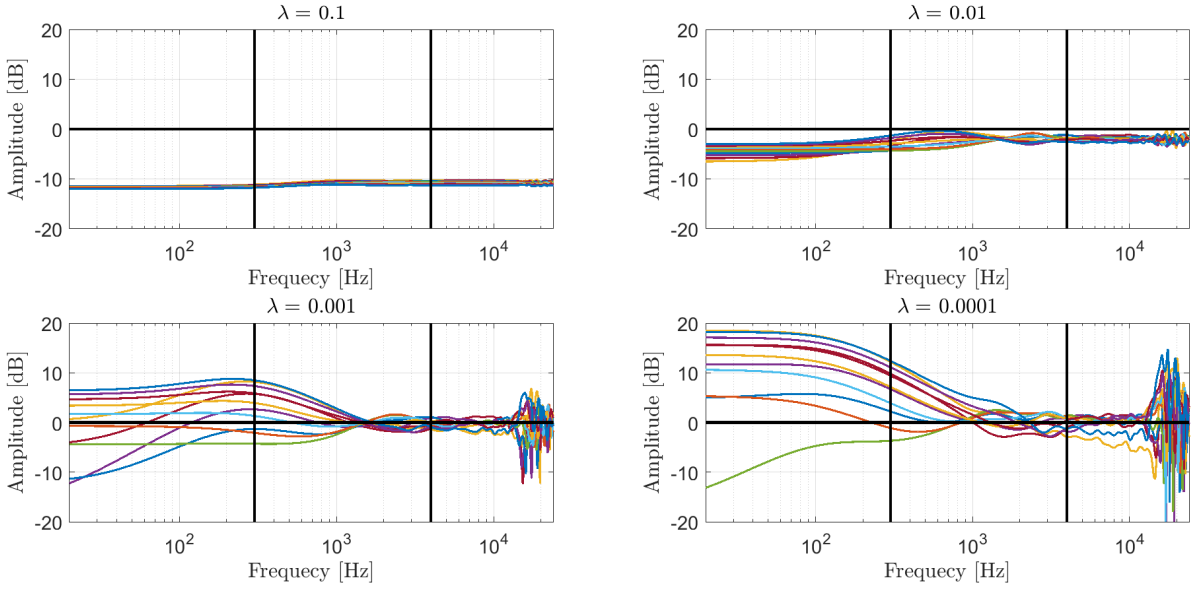


Figure 5.20: The amplitude response of the RSES6 filters for **setup 1**, with different values of λ.

Here it can be seen that for $\lambda = 0.001$ and $\lambda = 0.0001$ the filters are trying to boost, the frequencies within the region of interest. This can cause the filtered signal to clip, which is undesired.

Setup 2

Like in **setup 1**, different values of λ have been tested for **setup 2**, in figure 5.21, different RSES settings can be seen with different lambda settings. In figures 5.22 and 5.23, the filters in

frequency can be seen for the different values of λ. Again the reference and the RSES6 setting is shown.

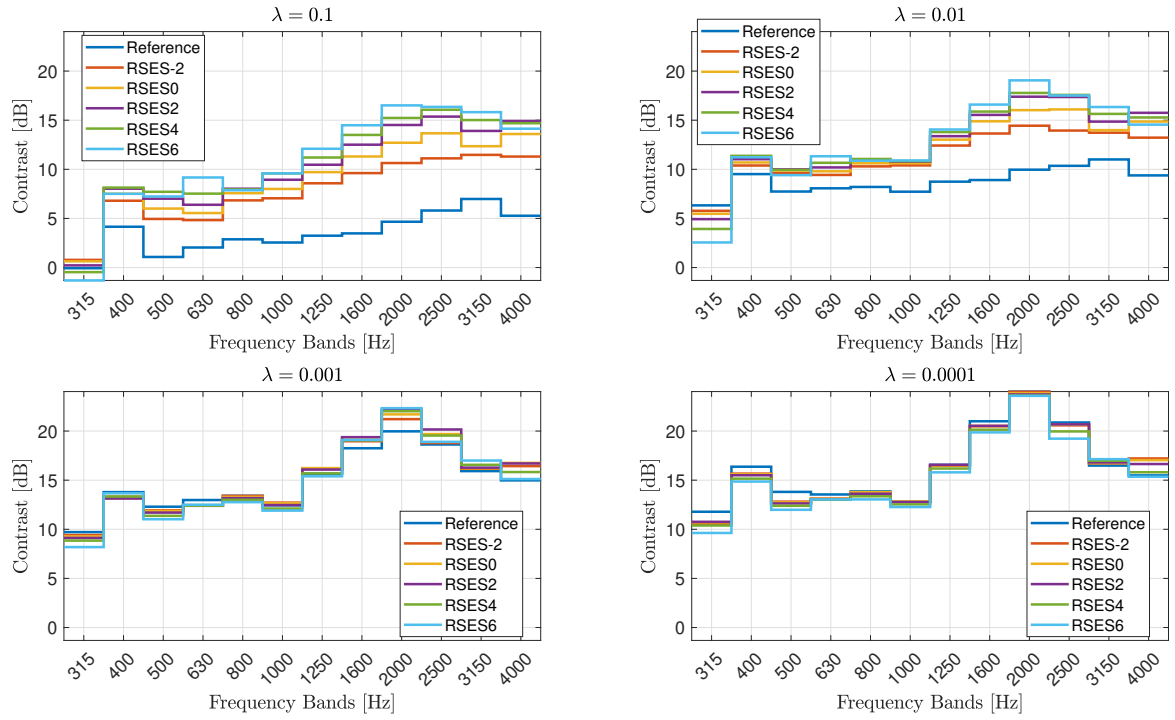


Figure 5.21: Contrast between the bright and dark control points, and the effect of using different values of λ in **setup 2**.

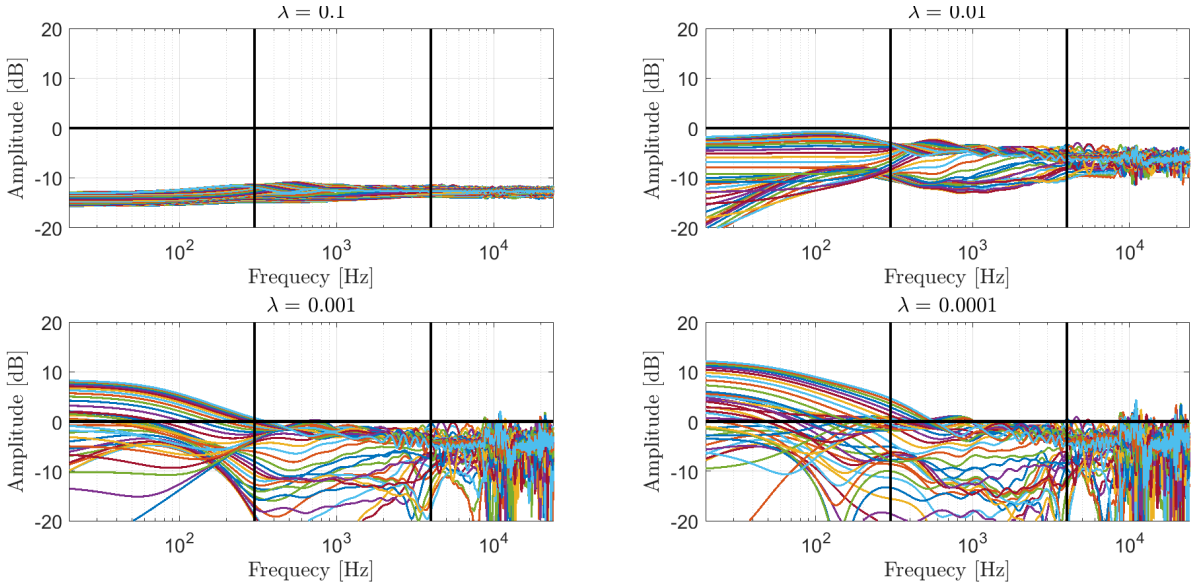


Figure 5.22: The amplitude response of the reference filters for **setup 2**, with different values of λ.

Similar to **setup 1**, it can be seen that for $\lambda = 0.001$ there is a slight boost at only a few frequencies, where for $\lambda = 0.0001$ the filters are trying to boost a lot of the frequencies within

the region of interest.

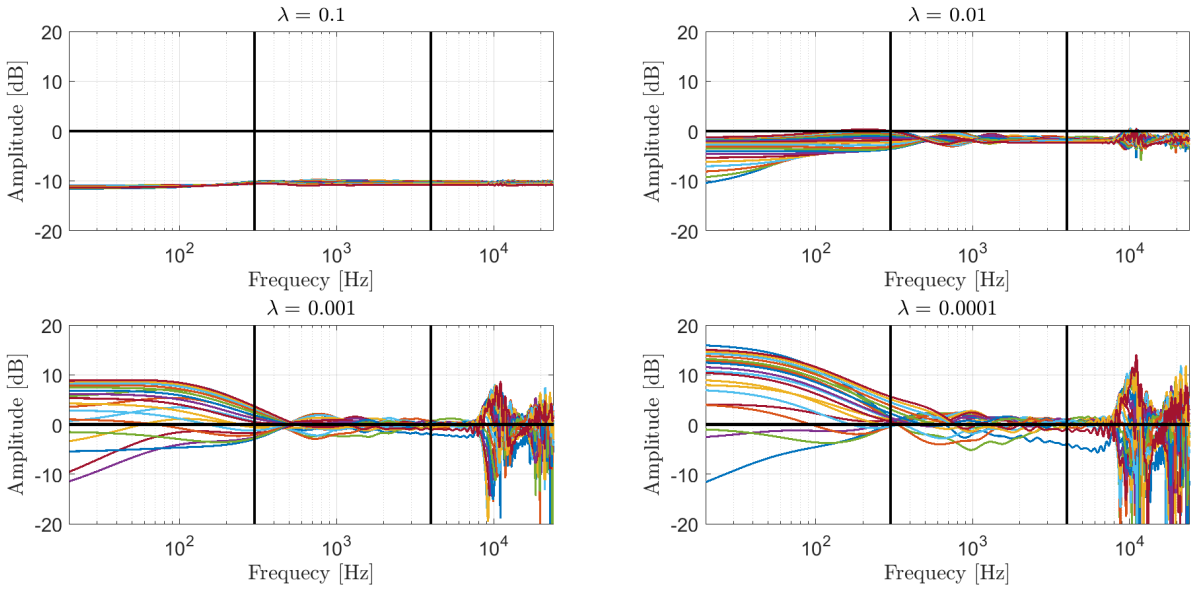


Figure 5.23: The amplitude response of the RSES6 filters for **setup 2**, with different values of λ.

In figure 5.23 that when using the RSES6 model, values of λ below 0.01 introduce boost within the entire region of interest.

5.4.1 Summary

It can be been in figure 5.18 and 5.21, that when decreasing the value of λ the ray space subdivision model show less increase in contrast compared to the reference model. When λ becomes very low, there is close to no difference using the ray space subdivision model or not. However if the λ value gets too low, this can cause the filters to boost specific frequencies, this is however very dependent on the position of the bright and the dark zone. Using the filters with the boosted frequencies can cause the filtered signals to clip. In order to avoid this, the values $\lambda = 0.1$ and $\lambda = 0.01$ have been chosen for following measurements.

5.5 Measurements Performed in a Real Room

In this section, the final validation test of the simulation model will be presented. The measurements are made by recording the sound pressure in the bright and the dark zone while playing a logarithmic chirp from 300 Hz to 24 kHz which have been filtered with the filters obtained from the optimization to the specific setup. The measuring setup is the same as described in section 5.1 where a full description of the used equipment can be found in appendix C.

All the measurements are made using the point source model, since the results from the simulations done with the measured impulse responses showed that the different ATF-models gave close to the same results. In the previous chapter, the results showed that the point source, piston model and measured RIR all had close to identical performance. However, the point source model was used for these tests being the least complex of the three, and does not require any measurements in order to do the optimization.

The two different setups are once again the two sub-optimal setups which can be seen in figure 5.24.

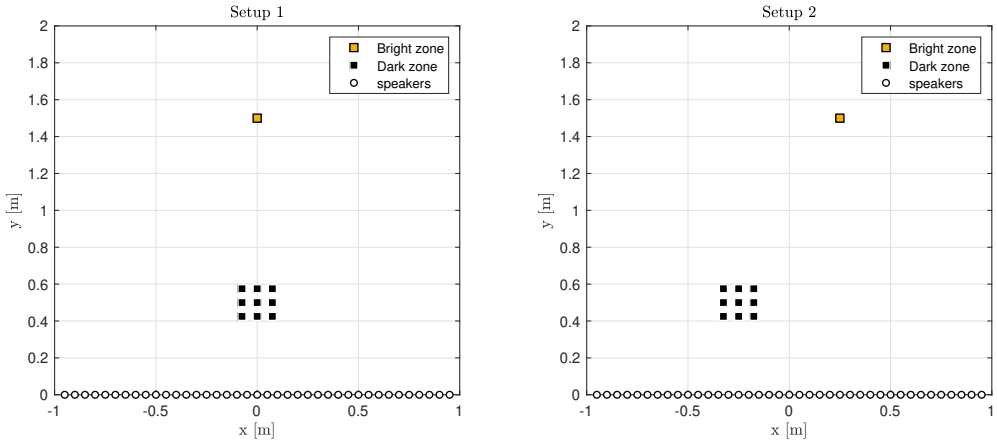


Figure 5.24: Plot of the control points positions relative to the loudspeaker array.

5.5.1 Results

In each of the setups, different values of λ as well as different RSES settings have been used in order to optimize the filters. The results section will be split up into two smaller subsections, one for each setup. These will contain figures showing both the contrast and the MSE for the different RSES settings. The MSE is calculated for each frequency, by first making a deconvolution of the played signal and the recorded signal. This will give the transfer functions of both the ATF and the filters for that specific recording, where the MSE is then calculated as described earlier in section 5.3.

Setup 1

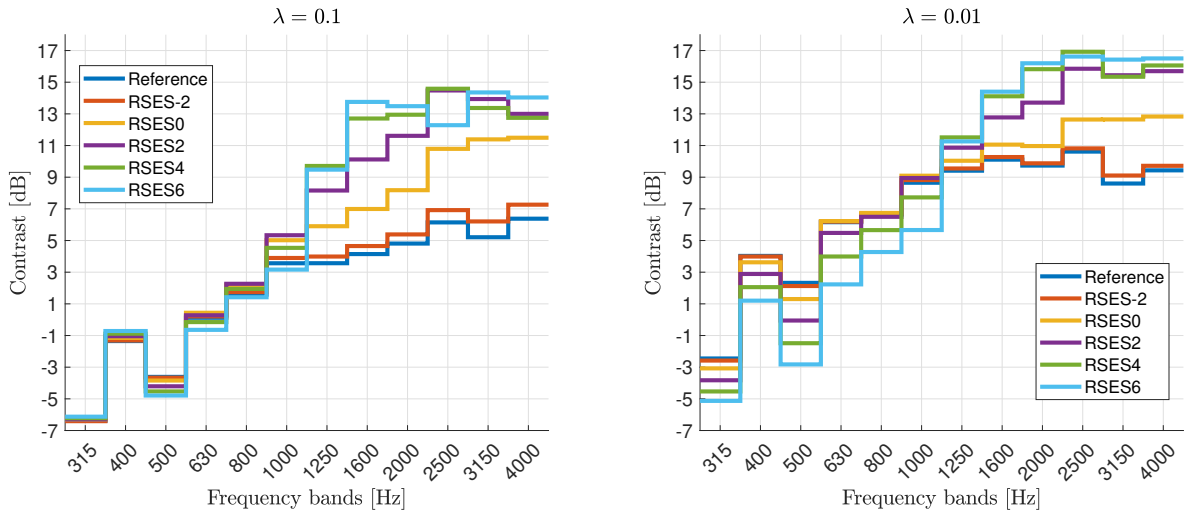


Figure 5.25: Measured contrast between the two zone in **setup 1** with different RSES settings. The bright control point is placed in $[0, 1.5]$ and the dark control points have center in $[0, 0.5]$.

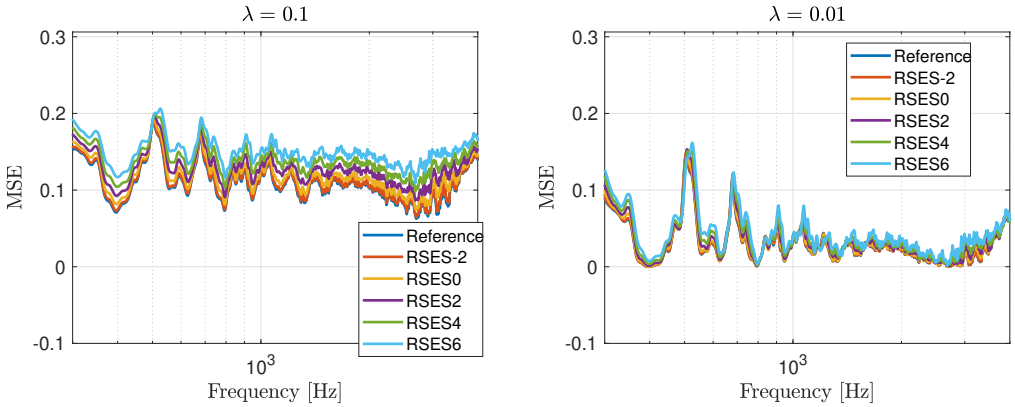


Figure 5.26: Calculated MSE using the measured data for **setup 1**.

In 5.25 it can be seen that for $\lambda = 0.1$ close to 9 dB increase in contrast is achieved, in the 3150 Hz frequency band, compared to the original wPM algorithm. In general it is seen that in the higher frequency bands more contrast is achieved compared to the lower bands where the contrast in general shows close to the same for all of the different settings. With $\lambda = 0.01$, in general the measurements show a smaller increase in contrast, and even a decrease in contrast at the lower frequencies when using the RSES-model. At the higher frequencies up to 7 dB increase in contrast is achieved in the 3150 Hz frequency band. Like in the previous section the MSE show very similar results as the simulation, where the MSE increases with the amount of loudspeakers turned off, this is seen in figure 5.26, However it can be seen that for $\lambda = 0.01$ the changes of the MSE when turning off additional speakers is very small.

Setup 2

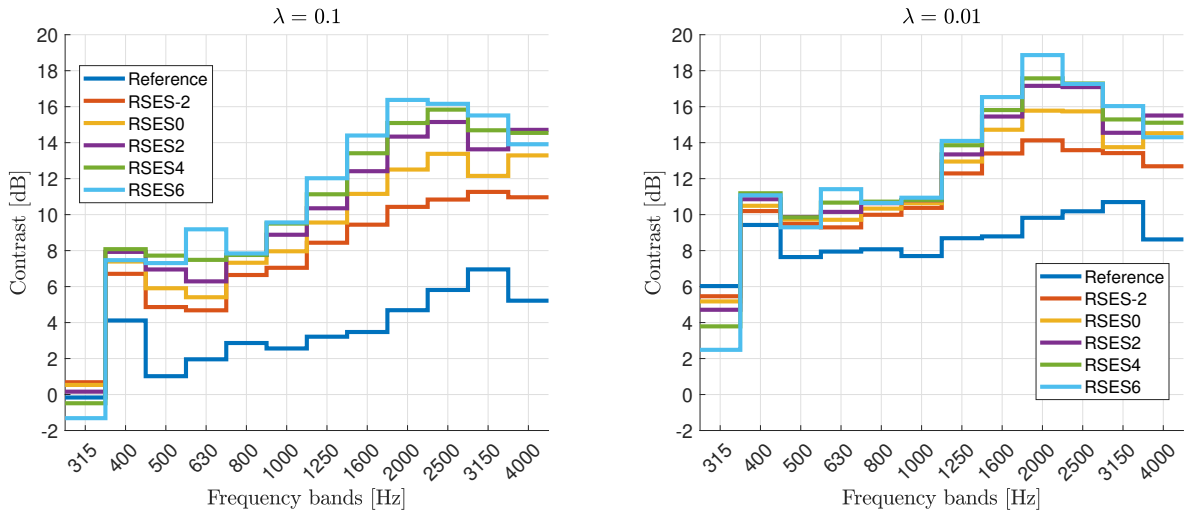


Figure 5.27: Measured contrast between the two zone in **setup 2** with different RSES settings. The bright control point is placed in $[0.25, 1.5]$ and the dark control points have center in $[-0.25, 0.5]$.

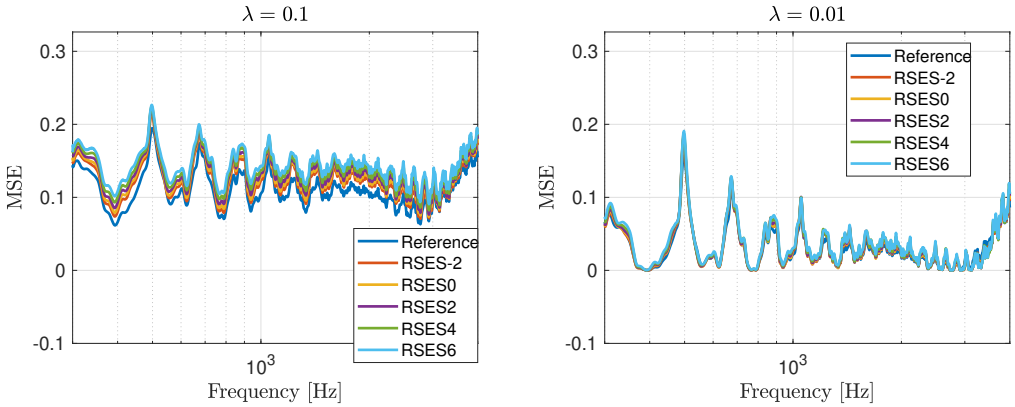


Figure 5.28: Calculated MSE using the measured data for **setup 2**.

In 5.27 it can be seen that for $\lambda = 0.1$ close to 12 dB increase in contrast is achieved in the 1600 Hz frequency band. In general compared to **setup 1**, **setup 2** achieves an increase in contrast throughout all of the frequency bands, except for the first one. For $\lambda = 0.01$ it is again seen that a smaller increase is achieved, here up to 9 dB is achieved in the 2000 Hz frequency band, but overall the increase is smaller, again the lowest frequency bands, does not increase from the RSES addition. Like in **setup 1** it can be seen that MSE is very similar for $\lambda = 0.1$ However for $\lambda = 0.01$ it can be seen that the there are barely any changes in the MSE, when turning off additional loudspeakers.

5.5.2 Informal Listening Test of the Reproduction of the Desired Signal

In order to evaluate the sound quality of the reproduced signal, a listening test is required. However as a large scale listening test is out of the scope for this project, an informal listening

test have been done by the authors. The main purpose of this test is to determine if applying the RSES-model alters the signal, and if this is the case how significant this change is.

The results from the informal listening tests, are that between the reference signal of the wPM algorithm, and in the cases where the RSES-model have been applied. The authors heard close to no difference in the reproduced signals, except for a very small change in the overall amplitude between the reference signal and with the RSES6-model applied with $\lambda = 0.1$. The slight reduction in the amplitude, did not feel disruptive to the listening experience. However if this were to be in a dynamic system, sudden changes in the amplitude could result in a worse listening experience. Spectrograms of the recorded signals have been made, where it can be seen that corresponding to the results of the listening test, a reduction in amplitude is the result of the RSES6 extension. This can be seen in figure 5.29, where the reference recording and the RSES6 recording have been shown, for **setup 2** and with $\lambda = 0.1$, the rest of the spectrograms can be seen in appendix F.

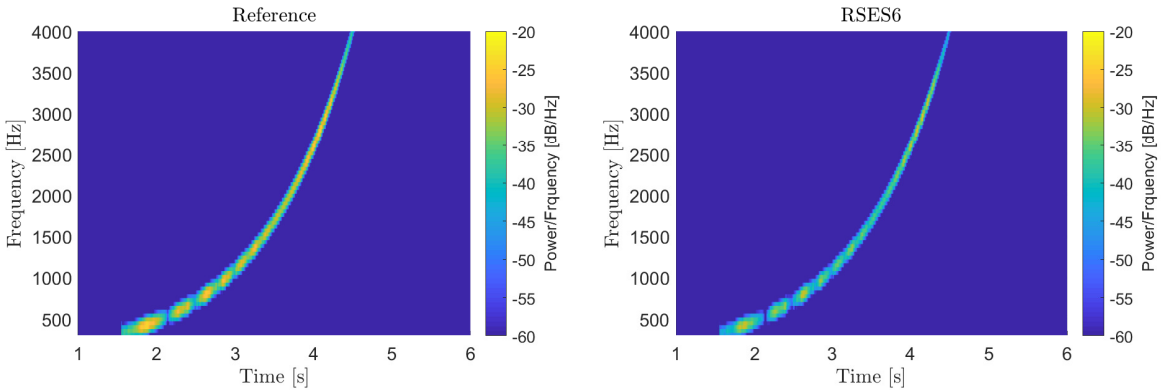


Figure 5.29: Spectrogram of the recorded signal, shown for the reference signal and with RSES6.

5.5.3 Uniformity of the Sound Pressure in the Bright Zone

Another important part of the sound experience in the sound zone, is uniformity of the sound pressure. Only one control point, in the bright zone, is used in the optimization, thus it could be interesting to look at the surrounding points. The results for **setup 1** can be seen in figure 5.30. The points that have been looked at are placed at: center [0, 1.5] (control point), left [0.175, 1.5] and right [-0.075, 1.5]. The results for for **setup 2** can be seen in figure 5.31, the points that have been looked at are placed at: center [0.25, 1.5] (control point), left [0.175, 1.5] and right [0.325, 1.5].

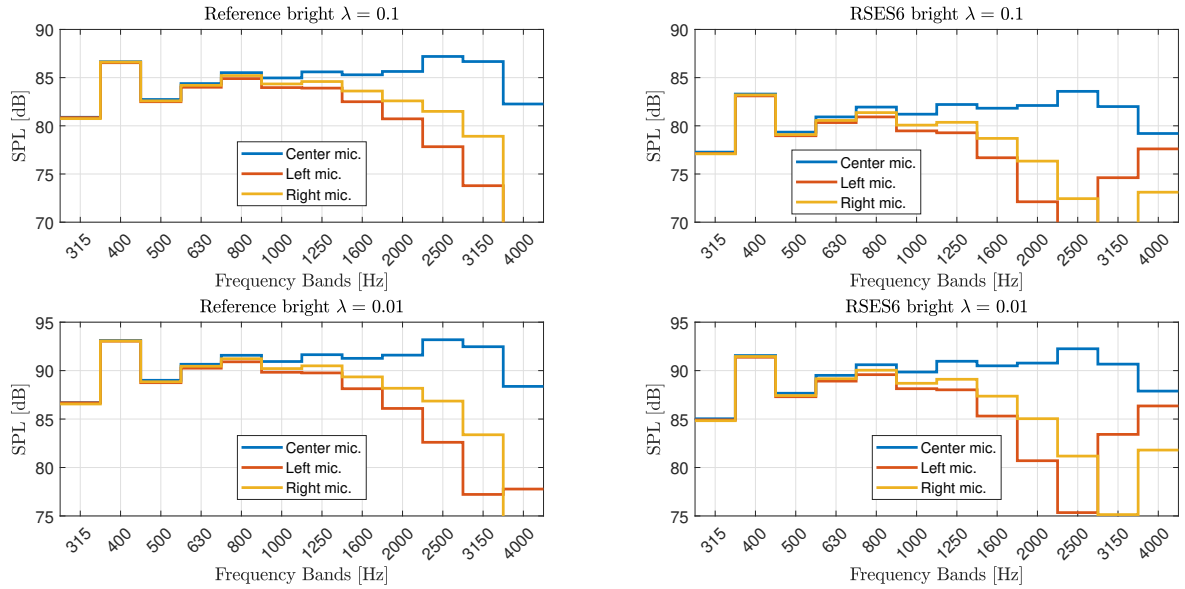


Figure 5.30: Sound pressure level in points surrounding the bright zone control point.

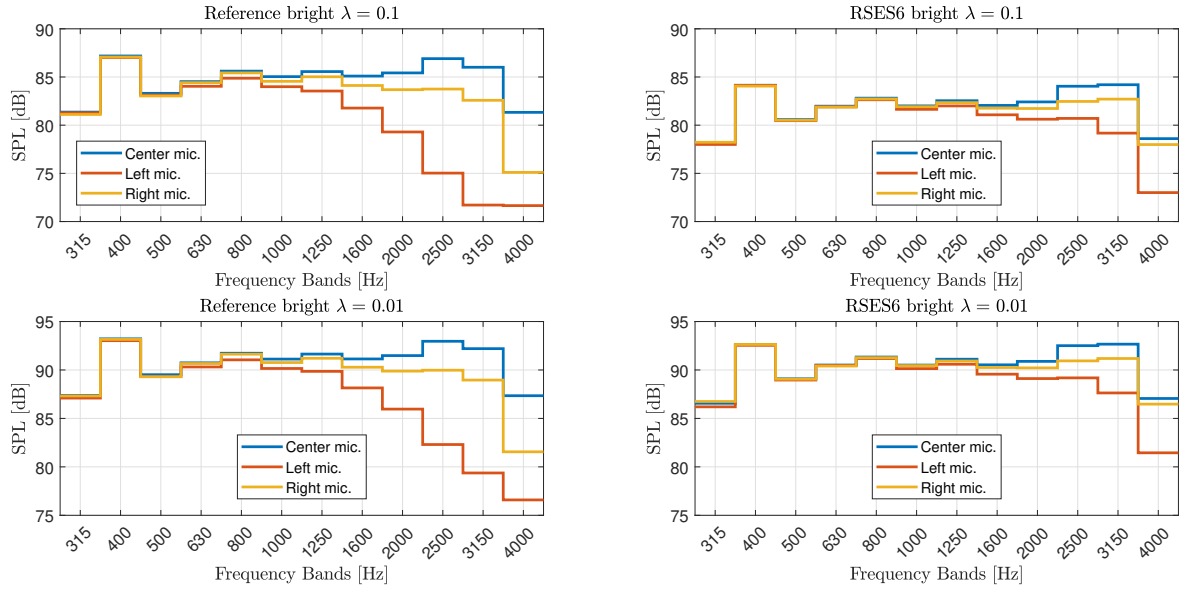


Figure 5.31: Sound pressure level in points surrounding the bright zone control point.

In figure 5.31 and figure 5.31 it can be seen that when using the RSES6 setting in **setup 1** a bigger spread in the sound pressure levels is seen around the frequency bands from 1000 - 3150 Hz, but a decrease in the 4 kHz frequency band. For **setup 2** the sound field in the measured point becomes more uniform.

5.6 Additional Test With Compensation for Additional Boost in Filters

As it was seen in the former section, the amplitude of the measured signal was reduced when turning off loudspeakers. It was seen that for $\lambda = 0.1$ the reduction was more significant than with a value of $\lambda = 0.01$. Due to this it could be interesting to compare the different λ-values and see how much the amplitude is reduced. However for values of λ below 0.01, a boost in the filters were observed, therefore in order to compensate for this boost, the amplitude of the played signal is decreased, to ensure that overflow does not happen. As it was also seen earlier in section 5.4, the difference in contrast between the reference and the RSES settings, was reduced when using a smaller lambda value.

It can be seen in the former figures, that the filters optimized with $\lambda < 0.01$ provided gain exceeding 0 dB in the frequency area of interest. As mentioned earlier, this can cause the signal to clip. A way to compensate for this additional gain is simply to turn down the amplitude of the played signal. The results of this have been simulated, and can be seen in figure 5.32 (**setup 1**) and 5.33 (**setup 2**). The simulation have been made with the reference where all the loudspeakers are turned and with RSES6. The test signal have been lowered with the highest boost value for the filters created with a given λ in the said area, to compensate for the additional gain. The filters have then been convoluted with the measured impulse response for the bright zone control point, and the SPL in the zone, for each value of λ, can be seen in the an 1/3 octave band plot.

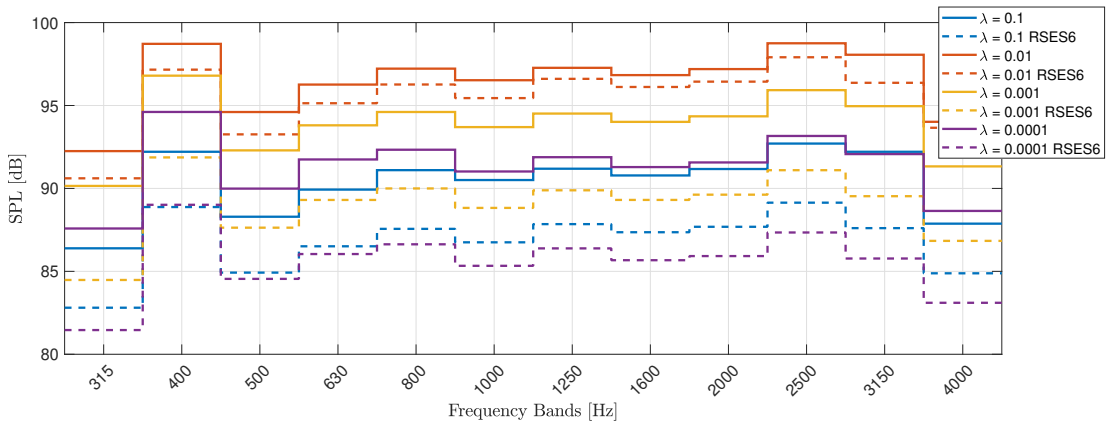


Figure 5.32: Simulations done with compensating for the additional gain provided by the filters created when using $\lambda = 0.001$ and $\lambda = 0.0001$. The simulations are done for **setup 1** and are for the reference and for the RSES6 setting. The plot show the 1/3 octave band plot of the sound pressure in the bright zone.

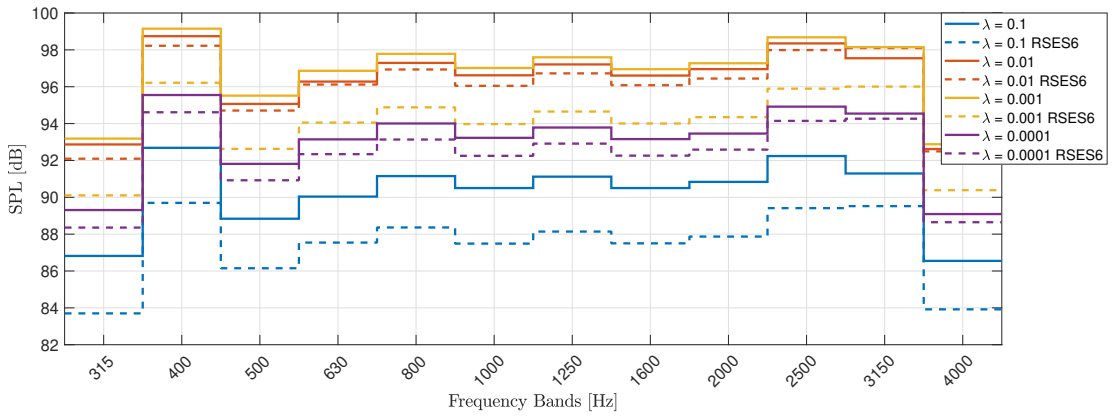


Figure 5.33: Simulations done with compensating for the additional gain provided by the filters created when using $\lambda = 0.001$ and $\lambda = 0.0001$. The simulations are done for **setup 2** and are for the reference and for the RSES6 setting. The plot show the 1/3 octave band plot of the sound pressure in the bright zone..

It can be seen in figures 5.32 and 5.33 that when compensating for the additional boost at the lower frequencies of the test signal, the sound pressure level is decreased in the bright zone. Thus the loudspeaker array can not be used to its full potential. As the value of λ gets low, it have been showed that using the ray space subdivisions model, in some cases a similar performance can be achieved without the model with the tradeoff being a lower sound pressure level, in other cases, the RSES6 model provide a worse performance compared to that of a lower values of λ. The results is very dependent on the specific setup.

5.6.1 Test of the Uniformity When Using a Low Regularization Parameter λ

Earlier, in section 5.5, it have been showed that when using the RSES-model, the sound pressure in the bright zone becomes more uniform when looking at point around the control point. Since lowering the value of λ have showed close to similar performance, in order to achieve contrast, with and without the RSES-model, it could be interesting to see how it perform when looking at the uniformity without the RSES-model. In the following plots, simulations have been made using the measured impulse responses used in chapter 5. In figure 5.34 the simulations are made with $\lambda = 0.1$ and $\lambda = 0.01$ using the RSES6 for **setup 1**, which have showed the best performance, in general, when looking at the uniformity of the bright zone[1]. This figure can then be compared with figure 5.35 which are simulated for **setup 1** using $\lambda = 0.001$ and $\lambda = 0.0001$. Figure 5.36 and 5.37 follows the same description as the two former figures, with the only different that they are simulated for **setup 2**.

[1]The plots showed are similar to the one showed in figure 5.30.

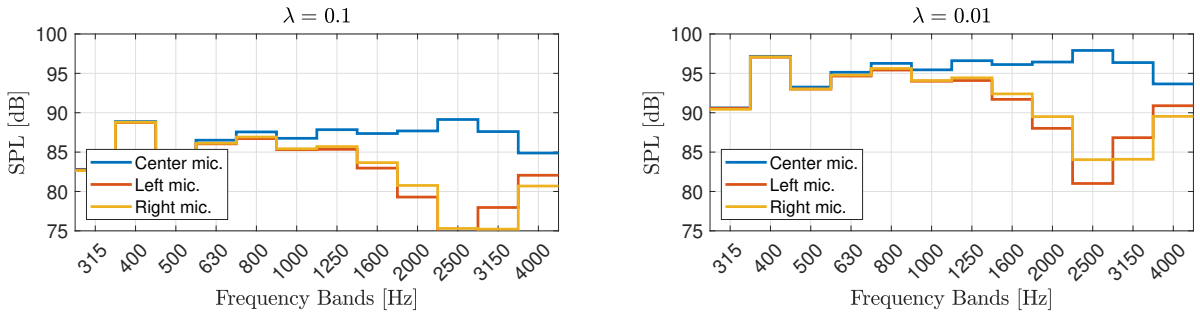


Figure 5.34: 1/3 octave band plot of the three different control point positions, showing the uniformity of the sound pressure in the bright zone. Results are for **setup 1** for $\lambda = 0.1$ and $\lambda = 0.01$ using RSES6.

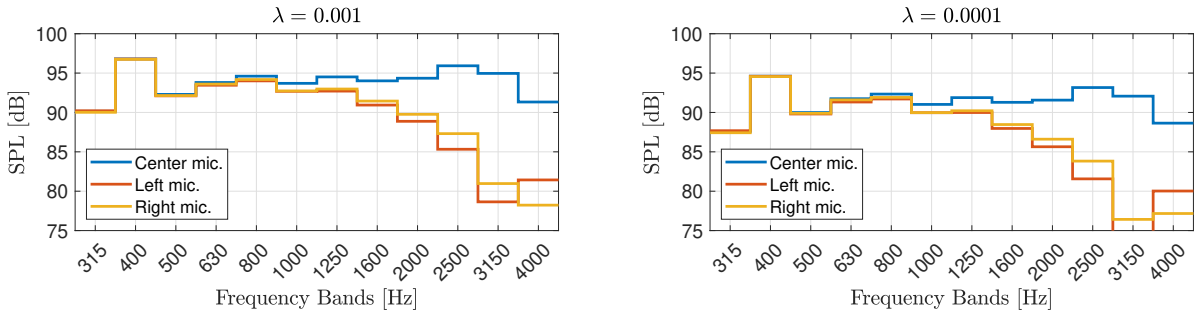


Figure 5.35: 1/3 octave band plot of the three different control point positions, showing the uniformity of the sound pressure in the bright zone. Results are for **setup 1** for $\lambda = 0.001$ and $\lambda = 0.0001$ without RSES.

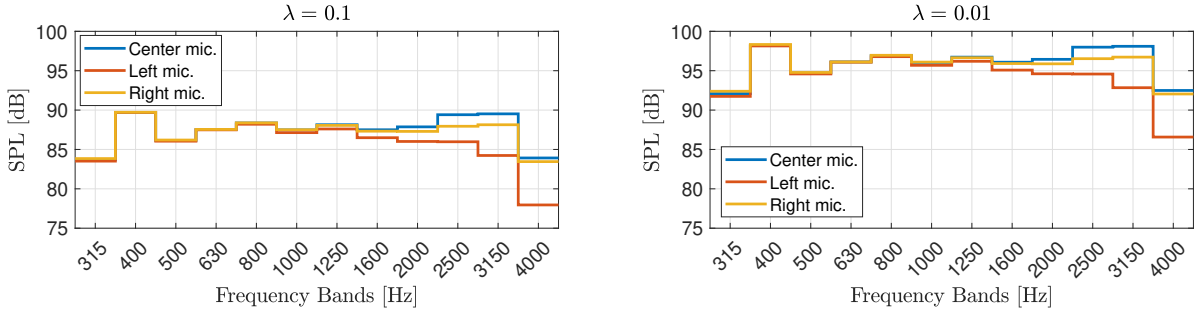


Figure 5.36: 1/3 octave band plot of the three different control point positions, showing the uniformity of the sound pressure in the bright zone. Results are for **setup 2** for $\lambda = 0.1$ and $\lambda = 0.01$ using RSES6.

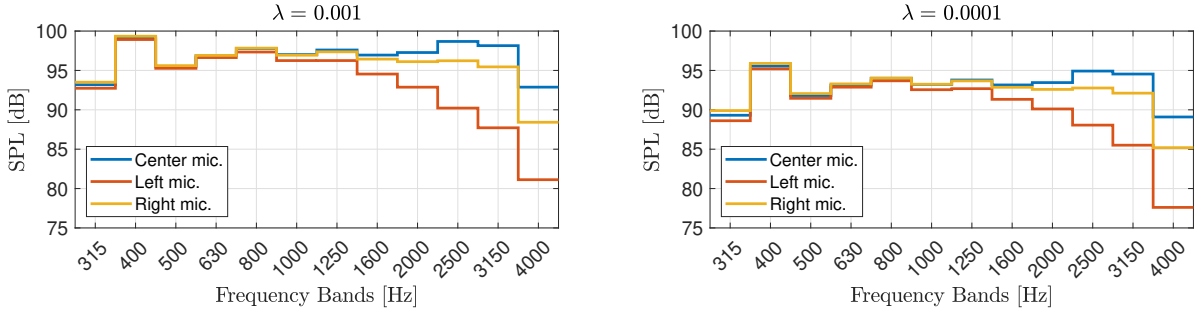


Figure 5.37: 1/3 octave band plot of the three different control point positions, showing the uniformity of the sound pressure in the bright zone. Results are for **setup 2** for $\lambda = 0.001$ and $\lambda = 0.0001$ without RSES.

The results showed in the figures 5.34 - 5.37 showed that using the RSES6-model improved the uniformity of the bright zone, when comparing $\lambda = 0.1$ and $\lambda = 0.01$ with RSES6 applied, compared to $\lambda = 0.001$ and $\lambda = 0.0001$ without the RSES-model. However it can be seen that for $\lambda = 0.1$ a lower sound pressure level compared to the other values of λ is achieved.

However, if using the RSES-model with $\lambda = 0.001$ and $\lambda = 0.0001$ similar uniformity can be achieved. However the tradeoff being a decrease in the sound pressure level, due to the compensation of the gain in the filters. This is illustrated in figure 5.38.

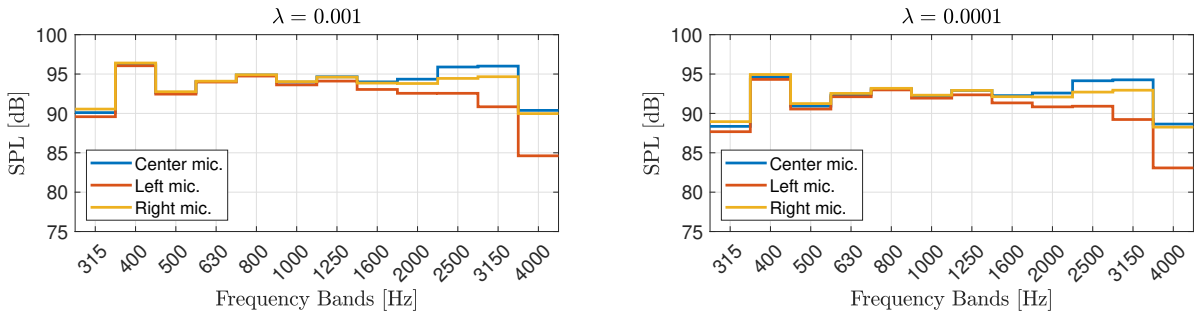


Figure 5.38: 1/3 octave band plot of the three different control point positions, showing the uniformity of the sound pressure in the bright zone. Results are for **setup 2** for $\lambda = 0.001$ and $\lambda = 0.0001$ RSES6.

6 | Discussion

In this chapter the discussion is divided into smaller subsections discussing the choices made throughout the project, the results and thoughts on future works.

In multizone sound control the superposition principle is used, meaning that in a lot of cases the same loudspeakers are used to create different sound zones. This does mean that if the sound zones are placed sub-optimally, one of the sound zones might have the direct path to some of the loudspeakers blocked. Meaning that sound from the one sound zone have to travel through the other sound zone. It has been examined how turning of critical loudspeakers does alter the performance of the multizone sound control. Two subdivision models RSMS and RSES were made. With the RSMS-model only turning off the loudspeakers directly blocked by the one zone, and the RSES-model which turns off the entire side of the array. The RSES-model was chosen for further testing.

Results

Looking at the results, it shows that the performance of using the RSES-model with the wPM algorithm is very dependent on both the specific setup, but also the specific settings of the wPM algorithm. In order to test the RSES-model, two different sub-optimal sound zone positions have been used. **Setup 1**, where the sound zones are placed directly in front of each other, and **setup 2**, where the zones are placed in front of each other with an offset from the middle. **Setup 1** was chosen because it was assumed to be to the worst case scenario. **Setup 2** was chosen based on the living room scenario where the two sound zones were placed equivalent to two seating positions on a corner sofa.

The results showed a big improvement in contrast for the measured results using the RSES-model. For **setup 1** with $\lambda = 0.1$, the RSES-model showed an increase in contrast in the frequency bands from 1250 Hz to 4000 Hz, here an increase of up to 9 dB was achieved in the 3150 Hz frequency band. For $\lambda = 0.01$ a decrease in contrast was seen in the 315 Hz to 1000 Hz frequency band, and an increase in the 1250 Hz to 4000 Hz frequency bands. Here an increase of up to 7 dB was seen in the 3150 frequency band. For **setup 2** the results showed an increase in contrast for all frequencies except in the 315 Hz frequency band. In this setup up to 12 dB was achieved for $\lambda = 0.1$ in the 1600 Hz frequency band, and up to 9 dB was achieved for $\lambda = 0.01$ in the 2000 Hz frequency band. For both setups, an increase in the MSE was seen when turning off loudspeakers. The changes in MSE for $\lambda = 0.1$ is assumed to be a results of an decreasing amplitude when turning off loudspeakers. However when $\lambda = 0.01$ barely any change is seen. This was probably due to the loudspeakers getting enough energy to recreate the signal, which is also seen in the overall lower MSE for $\lambda = 0.01$. This is supported by the informal listening where the only audible changes in the measured signals perceived by the authors, seemed to be a very small decrease in the amplitude.

General for both setups, it was seen that close to no contrast was achieved at the lower

frequencies. The lower frequencies of the region of interest was chosen based on the length of the array. It would therefore be interesting to test how the length of the array affects the performance and where the actual limit for lower frequency is placed in different setups.

It was interesting to see the decrease in amplitude when choosing a larger value for λ. Because of this additional simulations were made where the increased gain in the filters, for values of λ lower than 0.01, have been compensated for by decreasing the signal amplitude accordingly to avoid overflow in the played signals. The simulations showed that for λ lower than 0.01, using the RSES-model have close to no effect, looking at the contrast. However, as it is desired to be able to use the array at full amplitude, comparisons of the amplitude of the compensated signals with the ones using the RSES-model have been made. These showed that for **setup 2**, the simulation with $\lambda = 0.001$ had close to the same sound pressure level compared to simulation with $\lambda = 0.01$ and with RSES6 applied. Where the highest contrast was achieved with $\lambda = 0.001$ without the RSES6 applied. However when looking at **setup 1** it was seen that the amplitude for the compensated signals were lower than the one where RSES6 had been applied, thus showing that the results of the RSES-model is very dependent on the placement of the zones. However as a tradeoff, the value of λ is also used as an regularization factor which help to stabilize the matrix inversion and it was shown that smaller values of λ introduced more ringing in the impulse response. A listening test is required to determine the impact of the ringing.

Additional investigations have been made in order to see the sound pressure levels in points surrounding the bright zone control point. Results showed that when using the RSES-model in **setup 2** the sound pressure level gets more uniform compared to measurements made without the RSES-model, and simulations done with values of λ below 0.01, also showed an improvement in uniformity when using the RSES-model. Where not much of an improvement is seen using **setup 1**.

Additional Models of the Acoustic Transfer Function

Due to the fact that the wPM algorithm uses the acoustic transfer functions (ATF) for the optimization of the filters, it was assumed that a more precise ATF would improve the performance. Therefore two additional ATF models were introduced: a simple circular piston model and a room reflection model (ISM model). However from a comparison between the simulation and the measurements the ISM model showed a worse performance than the other models. This can be due to the reflection coefficient which for the ISM model is the same for all frequencies, where this is not assumed to be the case in the actual room. This could cause the ISM model to introduce some reflections or apply a boost to some frequencies which was not present in the measurement room. It could be tested whether frequency dependent reflection coefficients fitted to the actual room, would improve the performance of the ISM-model. The piston model and point source model showed close to identical performance. This is due to the fact that both point source and the piston model, have a close to identical response up until the higher frequencies which were out of the region of interest. However in general all the ATF-models showed very similar results, it could be interesting to test how the improved frequency dependent ISM-model would perform in a reverberant room and could be tested, in order to come close to the actual living room scenario.

Dynamic Regularization Parameter λ

Lowering the regularization parameter λ increased the performance of the wPM algorithm in terms of the contrast, but with the cost of the filters boost certain frequencies, and therefore introduced a risk that the filtered signal could overflow. To avoid this boost with too low values of λ compensations were made by decreasing the signal. However most of the boosted frequencies was found in the lower end of frequencies in the region of interest, future works could include making the value of λ dynamic, and see how it would affect the performance of the wPM algorithm. The gain in the filters does however also depend of the specific setup. Therefore an investigation how much of the gain are caused by an increase in distance and how much of it that are caused by the specific setup could be made. Thus the dynamic value of λ should both be dependent on frequency as well as the specific setup.

Uniformity and the wPM-TV Algorithm

Another part of the listening experience in the sound zone is the uniformity of the sound pressure level within the zones. If the sound pressure is only optimized to the control point, and the sound pressure level decreases with only small movements of the head, the listening experience might get worse. It could therefore be interesting to see how the uniformity for the bright zone affects the listening experience. In addition, it could be investigate how to make the sound pressure level in the sound zone more uniform e.g. by having more control points in the bright zone, and see how this affects the creation of the sound zones.

It was seen in sections 5.5.2 and 5.6, that the RSES-model improved the uniformity of the sound in the bright zone, and it could therefore be interesting to test the RSES-model with the wPM-TV algorithm, as this does have a constraint to ensure uniformity in the dark zone, but not in the bright zone. As the wPM-TV algorithm uses steepest descent to calculate the filter coefficients, it is very time consuming, therefore it could also be interesting to see how the RSES-model affects the convergence rate, as removing loudspeakers reduces the amount of possible solutions.

Another way to reduce the amount of calculations, is to calculate filters coefficients for less frequencies. It was tested how using frequency steps higher than 1 Hz, affected the filters. The results showed that ripple was introduced into the frequencies between the calculated filter coefficients. Further tests are needed to see the impact of these ripples, and how it affects the performance of the beamforming.

7 | Conclusion

Throughout the project the purpose has been to improve the performance of beamforming used for personal sound zones in sub-optimal positions, by subdiving a loudspeaker array into smaller subarrays, formulated in the problem statement as:

Can the performance of beamforming be improved in sub-optimal positions by dividing the line-array into smaller subarrays and strategically removing specific loudspeakers?

In this project examination of the wPM algorithm have been done. This includes test of the different coefficients in the algorithm in order to determine their affect on the performance of the beamforming. Based on these tests, simulations of the algorithm have been made with the focus to test sub-optimal sound zone positions. It was found that an increased contrast of the sound pressure between the sound zones can be achieved when strategically turning off loudspeakers, for the tested setups. A method to determine which loudspeakers that should be turned off was made using the ray space transform. Two ray space subdivision models were made: ray space middle subdivision (RSMS) and ray space edge subdivision (RSES), here the RSES-model had to biggest increase in contrast and was used for further testing. It was shown that the optimal filters for $\lambda = 0.001$ and $\lambda = 0.0001$ introduced gain in the filters, and real measurements were performed with $\lambda = 0.1$ and $\lambda = 0.01$. The tests showed similar performance between the simulation model and the actual measurements. Two sub-optimal setups was tested:

Setup 1

A bright zone is placed in [0, 1.5] and dark zone with center in [0, 0.5] the results showed that for the lower frequencies of in the region of interest, close to no improvement in contrast was found. For $\lambda = 0.1$ the frequencies from 1250 Hz to 4000 Hz, showed an increase in contrast compared to the reference model wPM and the RSES-model. Here an increase of up to 9 dB was seen in the 3150 Hz frequency band. When looking at the MSE an increase was seen in correlation with the amount of loudspeakers turned off. This agreed with the informal listening test, where a slight decrease in the amplitude was detected. For $\lambda = 0.01$ a decrease in contrast was seen in the frequency bands from 315 Hz to 1000 Hz, and an increase in the bands from 1250 Hz to 4000 Hz. Here an increase of up to 7 dB was seen again at the 3150 Hz band. For the MSE close to no change was seen when turning off additional loudspeakers, and no changes in the signal was detected from the informal listening test. Both results of the listening test, corresponded to what was seen in a spectrogram of the recorded signal.

Setup 2

A bright zone is placed in [0.25,1.5] and dark zone with center in [-0.25,0.5] the results showed that for $\lambda = 0.1$, an increase in contrast compared to the reference without the RSES-model, was seen in all frequency bands except for the 315 Hz band, here an increase of up to 12 dB was achieved in the 1600 Hz frequency band. The MSE showed as for **setup 1**, an slight increase correlated to the amount of loudspeakers turned off. For $\lambda = 0.01$, the results showed again that an increase in contrast was seen in all frequency bands except for the 315 Hz band, here

an increase in contrast of up to 9 dB was achieved in the 2000 Hz frequency band. The MSE again showed that close to no change in MSE was introduced from turning off loudspeakers. The results from the informal listening test showed that a slight change in amplitude could be heard for $\lambda = 0.1$ where no changes were detected for $\lambda = 0.01$, this corresponded to the spectrograms made of the recordings.

To conclude on this project, the results show that in some specific setups it is possible to improve the performance of the beamforming in sub-optimal positions, by dividing the line-array into smaller subarrays. However tests showed that an improved performances is very dependent on both the setup and the settings of the wPM-algorithm.

Bibliography

[1] Betlehem T, Zhang W, Poletti M, Abhayapala T. Personal Sound Zones: Delivering interface-free audio to multiple listeners. IEEE Signal Processing Magazine. 2015 03;p. 81–91.

[2] Francombe J, Mason R, Dewhirst M, Bech S. A Model of Distraction in an Audio-on-Audio Interference Situation with Music Program Material. JOURNAL OF THE AUDIO ENGINEERING SOCIETY. 2015 02;p. 63–77.

[3] Simón Gálvez MF, Elliott SJ, Cheer J. Time Domain Optimization of Filters Used in a Loudspeaker Array for Personal Audio. IEEE/ACM Transactions on Audio, Speech, and Language Processing. 2015 Nov;p. 1869–1878.

[4] Molés-Cases V, Pinero G, Gonzalez A, de Diego M. Providing Spatial Control in Personal Sound Zones Using Graph Signal Processing; 2019. p. 1–5.

[5] Møller MB. Sound Zone Control inside Spatially Confined Regions in Acoustic Enclosures. [PhD thesis]; 2019.

[6] Christensen JJ, Hald J. Technical Review Beamforming [Technical documentation]. Brüel & Kjær; 2004.

[7] Druyvesteyn WF, Garas J. Personal sound. Journal of the Audio Engineering Society. 1997;p. 685–701.

[8] Elliott SJ, Cheer J, Choi J, Kim Y. Robustness and Regularization of Personal Audio Systems. IEEE Transactions on Audio, Speech, and Language Processing. 2012 Sep;p. 2123–2133.

[9] Kinsler LE, Frey AR, Coppens AB, Sanders JV. FUNDAMENTALS OF ACOUSTICS, 4TH ED. Wiley and Sons, Inc.; 2000. ISBN: 0-471-84789-5.

[10] Kuttruff H. Room Acoustics. 4th ed. Spon Press; 2000. ISBN: 0-203-18623-0; page 323.

[11] Johnson D. Constrained Optimization [Webpage]. Jeffrey SilvermanKevin DuhElizabeth GregoryDon JohnsonEileen KrauseMariyah PoonawalaMatthew, JeanesKyle Clarkson; 7. juli 2003. Available from: `https://cnx.org/contents/n4Q14bKz@2/Constrained-Optimization`.

[12] Kreyszig E, Kreyszig H, Norminton EJ. Advanced Engineering Mathematics. 10th ed. Wiley; 2011. Isbn: 0470458364.

[13] Betlehem T, Teal PD. A constrained optimization approach for multi-zone surround sound; 2011. p. 437–440.

[14] Chang J, Jacobsen F. Sound field control with a circular double-layer array of loudspeakers. Acoustical Society of America Journal. 2012;p. 4518–4525.

[15] Canclini A, Markovic D, Schneider M, Antonacci F, P Habets EA, Walther A, et al. A Weighted Least Squares Beam Shaping Technique for Sound Field Control; 2018. p. 6812–6816.

[16] Coleman P, Jackson P, Olik M, Møller M, Olsen M, Pedersen J. Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array. The Journal of the Acoustical Society of America. 2014 04;p. 1929.

[17] Bender EA, Williamson SG. Lists, Decisions and Graphs With an Introduction to Probability. [E-book]; 2010.

[18] Shuman DI, Narang SK, Frossard P, Ortega A, Vandergheynst P. The emerging field of signal processing on graphs: Extending high-dimensional data analysis to networks and other irregular domains. IEEE Signal Processing Magazine. 2013 May;p. 83–98.

[19] Boyd S, Vandenberghe L. Convex Optimization. Cambridge University Press; 2004. Isbn: 978-0-521-83378-3.

[20] Simón Gálvez M, Elliott S, Cheer J. The effect of reverberation on personal audio devices. The Journal of the Acoustical Society of America. 2014 05;p. 2654.

[21] Coleman P, Jackson PJB, Olik M, Møller M, Olsen M, Abildgaard Pedersen J. Acoustic contrast, planarity and robustness of sound zone methods using a circular loudspeaker array a). The Journal of the Acoustical Society of America. 2014;p. 1929–1940.

[22] Baykaner K, Coleman P, Mason R, Jackson P, Francombe J, Olik M, et al. The relationship between target quality and interference in sound zones. Journal of the Audio Engineering Society. 2015 1;p. 78–89.

[23] Rämö J, Bech S, Jensen S. Validating a real-time perceptual model predicting distraction caused by audio-on-audio interference. The Journal of the Acoustical Society of America. 2018;p. 153–163.

[24] Moore BCJ. An Introduction to the Psychology of Hearing. Koninklijke Brill NV; 2013.

[25] Møller M, Olsen M. Sound Zones: On Envelope Shaping of FIR Filters. International Congress on Sound and Vibration (ICSV). The International Institute of Acoustics and Vibration (IIAV); 2017. p. 613–620. 24th International Congress on Sound and Vibration, ICSV24, ICSV24 ; Conference date: 23-07-2017.

[26] Zölzer U. DAFX - Digital Audio Effects. 2nd ed. John Wiley & Sons, Ltd.; 2002. ISBN: 0-470-84604-6; page 5.

[27] Arfken GB, Weber HJ. Mathematical Methods for Physicists. 6th ed. Elsevier Academic Press; 2005. Isbn: 0-12-088584-0; page 910.

[28] Bianchi L, Antonacci F, Sarti A, Tubaro S. The Ray Space Transform: A New Framework for Wave Field Processing. IEEE Transactions on Signal Processing. 2016 Nov;p. 5696–5706.

[29] Bianchi L, D'Amelio F, Antonacci F, Sarti A, Tubaro S. A plenacoustic approach to acoustic signal extraction; 2015. p. 1–5.

[30] Markovic D, Antonacci F, Sarti A, Tubaro S. Soundfield Imaging in the Ray Space. IEEE Transactions on Audio, Speech, and Language Processing. 2013;p. 2493–2505.

[31] Rabenstein R, Spors S. Spatial Aliasing Artifacts Produced by Linear and Circular Loudspeaker Arrays used for Wave Field Synthesis; 2006. p. 1418–1431.

[32] Allen J, Berkley D. Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America. 1979 04;p. 943–950.

[33] Maekawa Z, Rindel J, Lord P. Environmental and Architectural Acoustics. 2nd ed. Spon Press; 2011. ISBN: 978-0-415-44900-7; page 103.

[34] Dansk Standard. Akustik - Måling af rumakutiske parametre - Del 1: Rum til optræden. Danish Standards Foundation; 2009.

[35] Dansk Standard. Akustik - Måling af rumakutiske parametre - Del 2: Efterklangstid i almindelige rum. Danish Standards Foundation; 2008.

[36] Chan IH. Swept Sine Chirps for Measuring Impulse Response. SRS Audio. 2010;p. 1–6.

[37] Müller S, Massarani P. Transfer-Function Measurement with Sweeps. J Audio Eng Soc. 2001;p. 443–471. Available from: `http://www.aes.org/e-lib/browse.cfm?elib=10189`.

[38] MathWorks. audioPlayerRecorder. MATLAB; R2020a. Available from: `https://se.mathworks.com/help/audio/ref/audioplayerrecorder-system-object.html`.

[39] Department of Acoustics AAU. Multi channel listening room. Aalborg University B5-108; Last checked: 03-05-2020. Available from: `http://acoustics.aau.dk/facilities/facframe.html`.

[40] GRAS. GRAS 40AZ 1/2" Prepolarized Free-field Microphone, Low Frequency [Datasheet]. G.R.A.S; Last checked: 30-05-2020. Available from: `https://www.gras.dk/products/product/ss_export/pdf2?product_id=152`.

[41] GRAS. GRAS 26CC 1/4" CCP Standard Preamplifier with SMB Connector [Datasheet]. G.R.A.S; Last checked: 30-05-2020. Available from: `https://www.gras.dk/products/preamplifiers-for-microphone-cartridge/constant-current-power-ccp/product/206-26cc`.

[42] MathWorks. Calibration factor for microphone. MATLAB; R2020a. Available from: `https://se.mathworks.com/help/audio/ref/calibratemicrophone.html#mw_d24fb451-d47e-4460-92b3-46308c71f174`.

A | MATLAB code: wPM-TV

Code example A.1: Implementation of wPM-TV in MATLAB.

```
% speaker settings
spkSize = 2; % in inches
spk_spacing = (spkSize*2.5)*1e-2; % place the speakers, default: as close as possible
number_of_speakers = 21;

% zone location vectors
pv_d = [-0.55,0.95; -0.50,0.95; -0.55,1.05; 0.50,1.05]; % control points in dark zone
pv_b = [0.5, 0.5]; % control point in bright zone

% algorithm settings
f = 1000; % frequency [Hz]
xi = 0.995;
lambda = 0.1;
beta = 0.2;
mu = 0.1;
number_of_iterations = 5000;
precision = 2^(-23);

% speaker location vectors
% always odd number of speaker, so that center can be in zero
if not(mod(floor(number_of_speakers),2))
    number_of_speakers = number_of_speakers-1;
else
    number_of_speakers;
end
q = -floor(number_of_speakers/2)*spk_spacing: ...
                                spk_spacing: ...
                                floor(number_of_speakers/2)*spk_spacing;
p = [q', zeros(number_of_speakers,1)]; % speakers locations with center in zero

for j = 1:length(pv_b(:,1))
    for i = 1:length(p)
        dist_b(j,i) = norm(pv_b(j,:) - p(i,:));
    end
end
for j = 1:length(pv_d(:,1))
    for i = 1:length(p)
        dist_d(j,i) = norm(pv_d(j,:) - p(i,:));
    end
end

Mb = length(pv_b(:,1)); % control points in bright
Md = length(pv_d(:,1)); % control points in dark
db = ones(length(pv_b(:,1))); % desired signal in bright
```

```
% ----------------------------------TV-------------------------------------
% calculating sigma
for i = 1:Md
    for j = i+1:Md
        sigma(i) = norm(pv_d(i,:) - pv_d(j,:));
    end
end
sigma = 2/(Md*(Md-1))*sum(sigma,'all');

% calculating A-matrix
for i = 1:Md
    for j = 1:Md
        A(i,j) = exp(- (norm(pv_d(i,:) - pv_d(j,:))^2/(2*sigma^2)) );
    end
end
if not(isequal(A,transpose(A))) % A = A^T
    disp('something went wrong... A is not equal to A^T')
end

% calculating D and L
D = eye(size(A)).*sum(A,2);
L = D - A;

% creating transfer functions
c = 343; % speed of sound [m/s]
h = @(H,omega) exp(-1j*omega*H/c)./(4*pi*H+1); % transfer function
omega = 2*pi*f;
Hb = h(dist_b,omega); % transfer function matrix, bright
Hb(find(Hb >1)) = 1; % to avoid stability when creating surface plot
Hd = h(dist_d,omega); % transfer function matrix, dark
Hd(find(Hd>1))=1;


% initialize g in steepest descent
% The optimal solution (gradient)Jwpm(g)= 0
gwpm = (k/Mb * (Hb'*Hb) + (1-k)/Md * (Hd'*Hd) + lambda*eye(length(Hd'*Hd)))\(k/Mb * Hb'*db);
g_new = gwpm;

for iteration = 1:number_of_iterations % steepest descent algorithm
    g_old = g_new;
    for m = 1:Md
        fd(m,1) = sqrt(g_old'*Hd(m,:)'*Hd(m,:)*g_old);
        J(m,:) = (1/fd(m))*((Hd(m,:)'*Hd(m,:))*g_old);
    end
    g_new = g_old - mu*(2*(k/Mb * (Hb'*Hb*g_old - Hb'*db) ...
                    + (1-k)/Md * (Hd'*Hd)*g_old + lambda*g_old) + 2*beta*transpose(J)*L*fd);
    if max(abs(g_new - g_old)) < precision
        break
    end
end
gtv = g_new
```

B | MATLAB code: ISM

Code example B.1: Implementation of ISM in MATLAB.

```
function [rir, rir_cell] = ISM_RIR_multi(Fs, pv_b, pv_d, speakers, zHeight, room, alpha, n)
% Function that creates the room impulse response (RIR) of the room using the image source
% method found in "Image method for efficiently simulating small-room acoustics"
% by Jont B. Allen and David A. Berkley, 1978/1979.
%
% The input are:
%          Fs       = sampling frequency [Hz]
%          pv_b     = bright zone control points location vector [m]
%          pv_d     = dark zone control points location vector [m]
%          speakers = loudspeaker coordinate vector [m]
%          zHeight  = height of the loudspeaker array and control points
%          room     = dimensions of the room, [x,y,z] [m]
%          alpha    = absorption coefficients for the different walls:
%                     [x1, x2, y1, y2, z1, z2], 0 < alpha <= 1
%          n        = The program will account for (2*n+1)^3 virtual sources

mic = [[pv_b;pv_d],ones(length(pv_b(:,1)) + length(pv_d(:,1)),1)*zHeight];
source = [speakers, ones(length(speakers(:,1)),1)*zHeight];

beta = -abs(sqrt(1-alpha)); % -abs(), to allow inversion of impulses

N=-n:1:n;

p = ((-1).^(N)*(-1) + 1)*0.5;
m = (N+0.5-0.5*(-1).^N)*0.5;
[mx,my,mz] = ndgrid(m, m, m); % create 3D matrix for reflections coefficients
[px,py,pz] = ndgrid(p,p,p);

t1 = datetime('now','Format','HH:mm:ss.SSS');
for jj = 1:length(mic(:,1))
    fprintf(['Calculating for control point ', num2str(jj),'/',num2str(length(mic(:,1)))])
    tic
    for ii = 1:length(source(:,1))
        Rp(ii,:,:) = [source(ii,1) - 2*p*source(ii,1) - mic(jj,1); ...
                      source(ii,2) - 2*p*source(ii,2) - mic(jj,2); ...
                      source(ii,3) - 2*p*source(ii,3) - mic(jj,3)];
    end

    for ii = 1:length(source(:,1))
        Rm(ii,:,:) = [2*m*room(1);...
                      2*m*room(2);...
                      2*m*room(3)];
    end
    Rpm = Rp + Rm;
    for ii = 1:length(source(:,1))
        [i,j,k] = ndgrid(squeeze(Rpm(ii,1,:))', squeeze(Rpm(ii,2,:))', squeeze(Rpm(ii,3,:))');
        dist(:,:,:,ii) = sqrt(i.^2 + j.^2 + k.^2);
    end
```

```
49
50      betaCal = (beta(1).^(abs(mx-px))).*(beta(2).^abs(mx)).*...
51                (beta(3).^(abs(my-py))).*(beta(4).^abs(my)).*...
52                (beta(5).^(abs(mz-pz))).*(beta(6).^abs(mz));
53
54      for ii = 1:length(source(:,1))
55          e(:,:,:,ii) = betaCal./(squeeze(4*pi*dist(:,:,:,ii)));
56      end
57
58      time = round(Fs*dist/343)+1; % makes all the distances to samples
59      for ii = 1:length(source(:,1))
60          temp_time = squeeze(time(:,:,:,ii));
61          temp_e = squeeze(e(:,:,:,ii));
62          rir_cell{jj,ii} = full(sparse(temp_time(:),1,temp_e(:)));
63      end
64      t = toc;
65      fprintf([' - time used: ',num2str(t),' s\n'])
66  end
67  t2 = datetime('now','Format','HH:mm:ss.SSS');
68  fprintf(['Total time used: ',char(t2 - t1),'\n'])
69
70  for jj = 1:length(mic(:,1))
71      for q = 1:length(source(:,1))
72          test = q + (length(source(:,1))*(jj-1));
73          temp = rir_cell{jj, q};
74          if q > 1 || jj > 1
75              if length(temp) < length(rir(:,1))
76                  temp(numel(rir(:,1))) = 0;
77              else
78                  rir(size(temp,1), size(rir,2)) = 0;
79              end
80          end
81          rir(:,jj,q) = temp;
82      end
83  end
84  end
```

C | Setup for Real Room Measurements

In this appendix, the measurement setup used for the measurements made in a real room is described. The description includes what hardware is used, how it is connected and how it interact with the software. Additionally, the calibration procedure of the microphones is described. A picture of the measurement setup can be seen in figure C.1 and C.2 for one of the microphone setups.

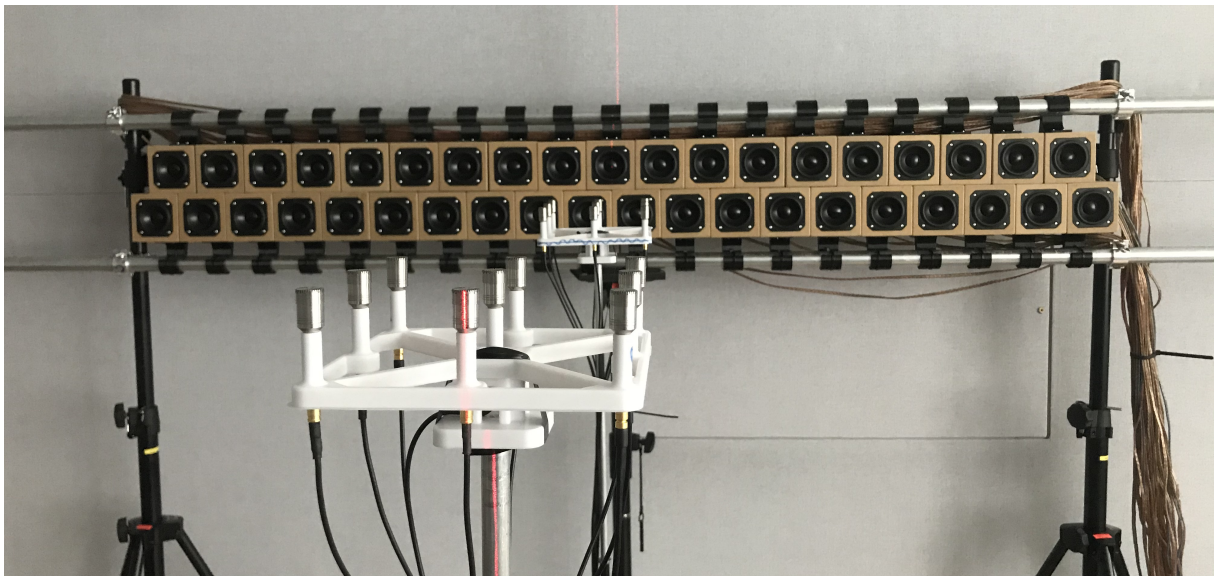


Figure C.1: Picture of one of the measurement setups.

Figure C.2: Picture of one of the measurement setups.

C.1 List of Equipment

Equipment	AAU No. or Serial No.	Type
RME Fireface UFX+	108249	Soundcard
RME Micstasy	86848	AD Converter
RME Micstasy	86847	AD Converter
RME M-16 DA	86844	DA Converter
RME M-16 DA	108243	DA Converter
RME M-16 DA	86844	DA Converter
7x150 W ICEPOWER amplifier	150ASH7, S/N 0001	Power amplifier
7x150 W ICEPOWER amplifier	150ASH7, S/N 0002	Power amplifier
7x150 W ICEPOWER amplifier	150ASH7, S/N 0003	Power amplifier
7x150 W ICEPOWER amplifier	150ASH7, S/N 0004	Power amplifier
7x150 W ICEPOWER amplifier	150ASH7, S/N 0005	Power amplifier
7x150 W ICEPOWER amplifier	150ASH7, S/N 0007	Power amplifier
GRAS 40AZ, w. 26CC preamp	78162	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78161	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78168	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78165	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78045	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78169	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78049	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78187	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78031	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78136	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78170	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78029	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78164	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78185	Measuring microphone
GRAS 40AZ, w. 26CC preamp	78171	Measuring Microphone
GRAS 40AZ, w. 26CC preamp	78167	Measuring Microphone
GRAS 40AZ, w. 26CC preamp	78041	Measuring Microphone
GRAS 40AZ, w. 26CC preamp	78163	Measuring Microphone
Microphone calibrator 94 dB ref. 20 μPa	78301	Brüel & Kjær - TYPE 4231
39 x 2 inch loudspeaker	-	Loudspeakers
3D-printed microphone array mount	-	array mount
MATLAB R2019b	-	Data processing software

Table C.1: List of equipment used for various tests.

C.2 Hardware/Software Connection

A diagram of the hardware connections can be seen illustrated in figure C.3.

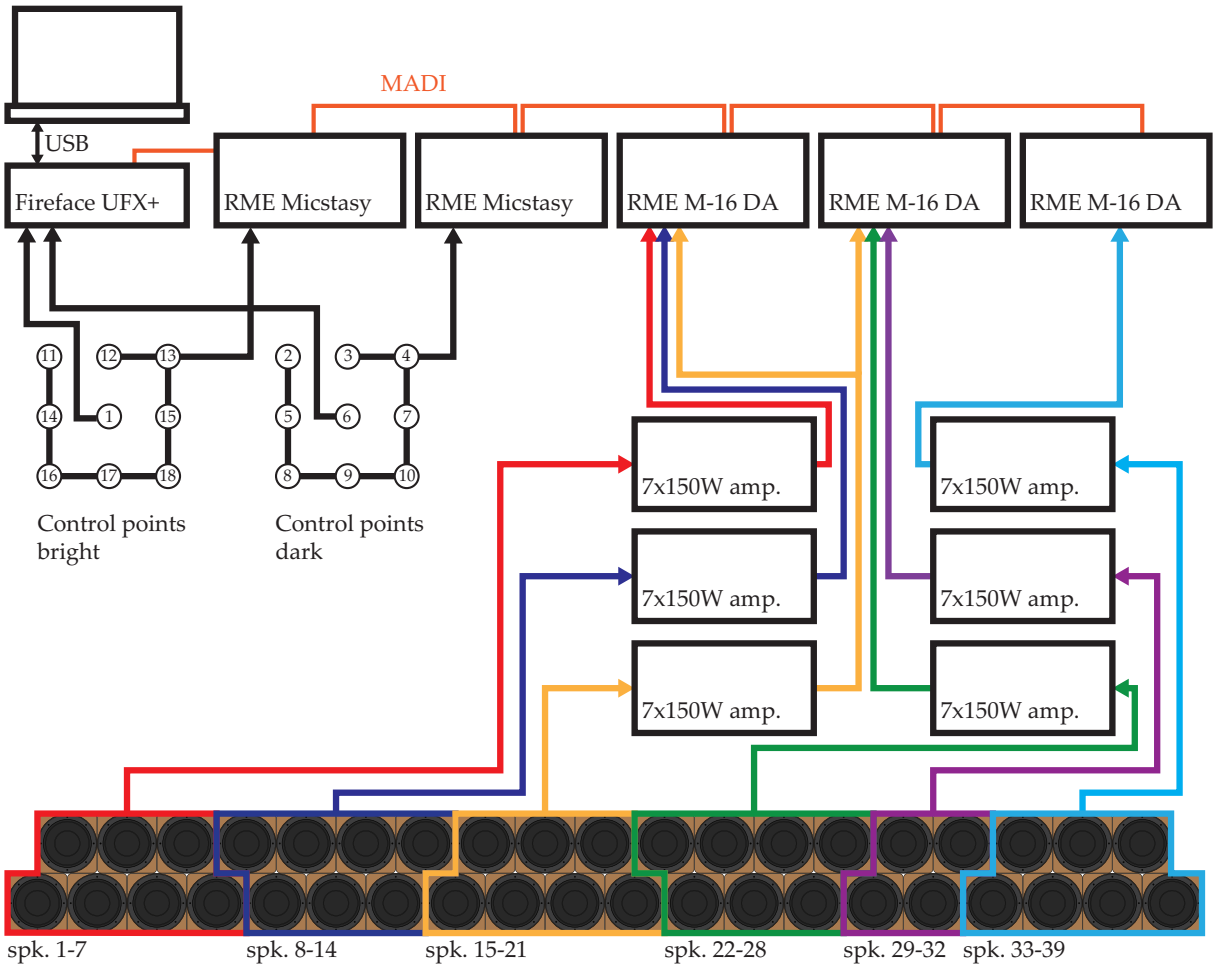


Figure C.3: Caption

The sound card is connected to the AD/DA-converters through MADI (Multichannel Audio Digital Interface). To control the inputs and outputs MATLAB is used, using the `audioPlayerRecorder()` object from the Audio Toolbox [38]. This object makes it possible to simultaneously play and record a signal.

C.3 Room Conditions

The room, B5-108 at Aalborg University, is a symmetric room and with a reverberation time at approximately 0.2 - 0.4 seconds, depending on the reflective/absorbing movable panels and the room conditions [39]. However, equipment and another test setup are present in the room, which can cause more reflections than normal. Panorama pictures of the room can be seen in figure C.4 and C.5.

Figure C.4: Panorama picture of the room.

Figure C.5: Panorama picture of the room.

The dimensions of the room are $[8.088, 7.346, 2.865]$ meters. A sketch of the room can be seen in figure C.6, where the placement of the setup also can be seen. The temperature in the room doing all the measurements is measured to $23 - 24$°C. For practical reasons, two persons were present in the room during the test of the personal sound zones.

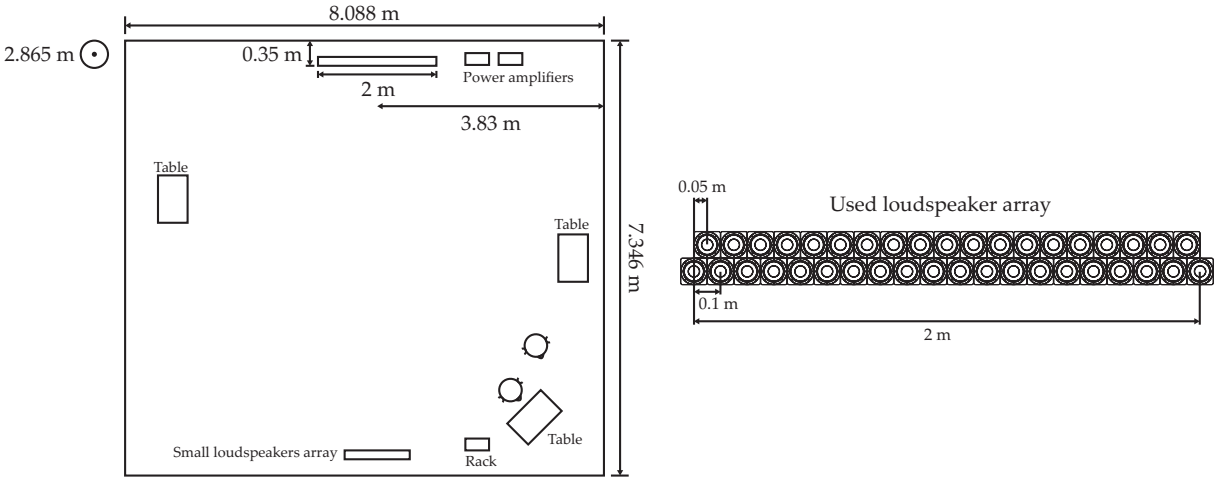


Figure C.6: Sketch of the room and the loudspeaker array.

C.4 The Microphones and Calibration Procedure

The microphones used are GRAS 40AZ [40] with a GRAS 26CC preamplifier [41]. The microphones are free-field microphones meaning that a low pass filter is applied at the higher frequencies, which will be equalized to have a flat amplitude response when the orientation of the microphone is 0° relative to the sound source. However, for practical reasons, the microphones are orientated 90° relative to the loudspeakers. Thus a the effect of the low pass filter may be expressed. The amplitude response of the microphone can be seen in figure C.7.

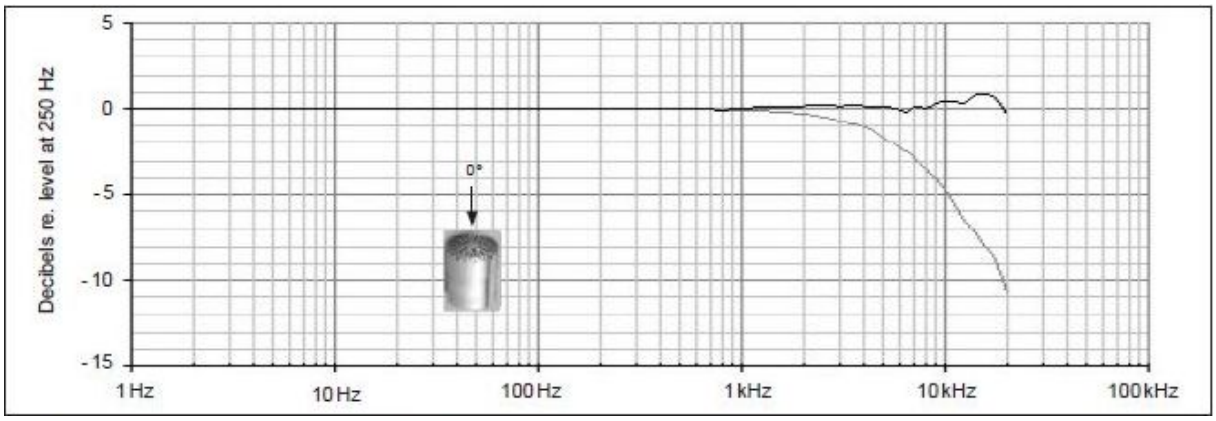


Figure C.7: Amplitude response of the GRAS 40AZ. The black line is the response when the orientation of the microphone is 0° relative to the sound source, and the grey line is the actual response of the microphone with the low pass filter [40].

To know the conversion through the system from the measured pressure at the microphone membrane to the electric signal and then into a digital signal, calibration of all the used microphones have been made. The calibrations are made with a microphone calibrator which produce 94 dB ref. 20 μPa. To get the calibration value equation (C.1) is used [42]:

$$\text{Calibration value} = \frac{10^{(\text{TrueSPL}-k)/20}}{\text{rms}(x)} \tag{C.1}$$

TrueSLP is equal to 94 dB, x is the recorded signal and k is 1 Pa relative to the pressure reference (ref. 20 μPa) in dB:

$$k = 20\log_{10}\left(\frac{1}{\text{pressure reference}}\right) \tag{C.2}$$

The calibration is as followed:

1. Turn on the microphone calibrator and place it on the a microphone
2. Record 5 seconds, and save the recording
3. Repeat the two previous steps until all microphone have been calibrated
4. Use equation (C.1) on all the recordings and make a vector with all the calibration values.

The calibration values can be seen in table C.2:

Microphone nr.	1	2	3	4	5	6	7	8	9
Calibration value	16.5028	24.6024	24.7324	26.0480	25.8360	17.0366	25.7078	23.9894	24.9291

Microphone nr.	10	11	12	13	14	15	16	17	18
Calibration value	24.1584	27.1675	26.8989	26.2796	25.3416	27.6757	26.9061	25.9265	22.6723

Table C.2: Calibration values for each microphone.

C.5 Noise Floor Measurement

To see when the microphones reach the noise floor, a test have been made. Simply a 10 seconds recording with all microphones have been made, the calibration values are multiplied to the recorded signals and an octave band analysis have been made. The results can be seen in figure C.8.

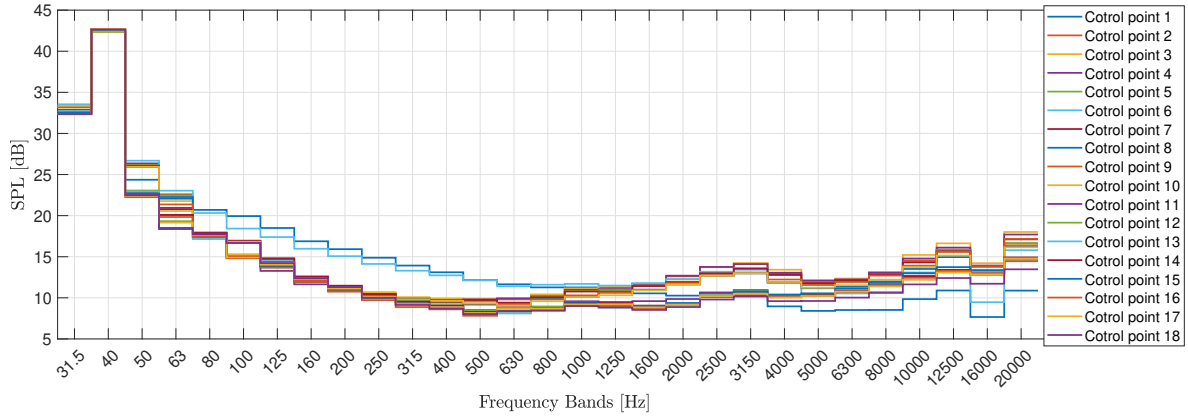


Figure C.8: Noise floor for each microphone showed in 1/3-octave bands.

D | Measuring Impulse Responses for the Control Points

The impulse responses of the transfer function between each loudspeaker and each control point are measured. These are created using the Swept Sine Method, which will be described, followed by the measuring procedure, data processing and results. The same procedure is done for each of the tested setups.

D.1 Swept Sine Method

To create the impulse responses for each loudspeaker to each control point, the Swept Sine Method is used. The method follows the structure illustrated in figure D.1.

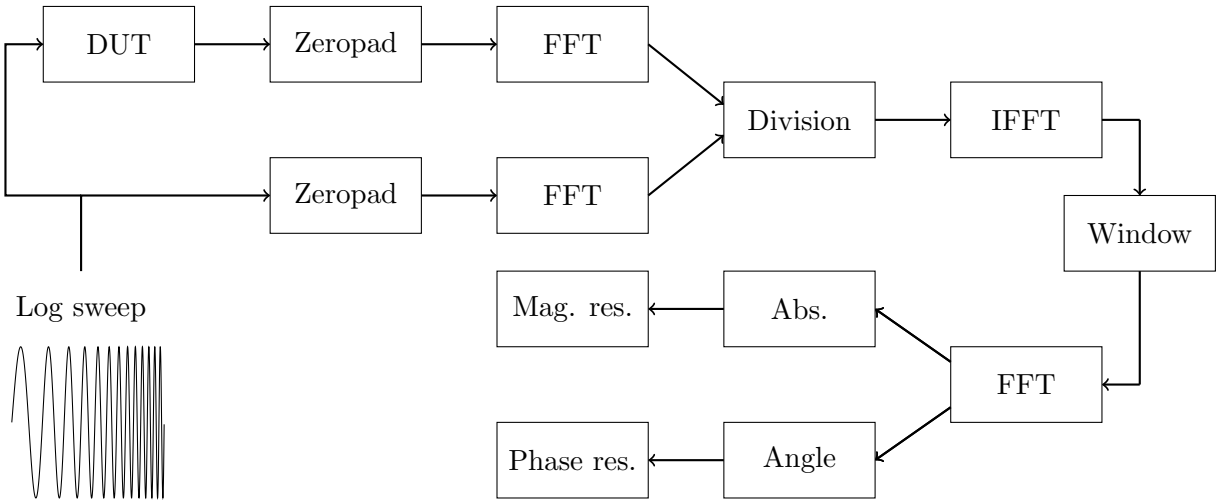


Figure D.1: Flow chart for Swept Sine Method [36]

A logarithmic chirp is generated, which is the reference signal. The output from the DUT (device under test), is the microphone recordings of the played signal. The excitation chirp and the output from the DUT, are then zero padded with n samples, where n is equal to the length of the signal, resulting in doubling of their previous sizes. Then an FFT is performed on the two signals and the recorded signal is then divided with the reference signal. Afterwards, an IFFT is used to obtain the impulse response, where the negative arrival times can be ignored by using a window that only takes the positive time arrivals [36, 37]. An FFT can be used in order to either get the magnitude response or the phase response of the impulse.

When using the Swept Sine Method, it can be advantageously to utilize its ability to suppress uncorrelated noise, which is assumed for the additional noise on the recordings. By repeating the measurements, the impulses can be averaged and will suppress the uncorrelated noise and a higher signal-to-noise ratio (SNR) will be achieved [37]. This can be seen simulated in figure

D.2, where the amplitude response of the impulses are plotted. The input signal is a logarithmic chirp from 20 Hz to 24 kHz and the signal from the DUT is the same chirp but with some random noise added. The figure show the effect of averaging over more recorded signal, which all have uncorrelated noise added.

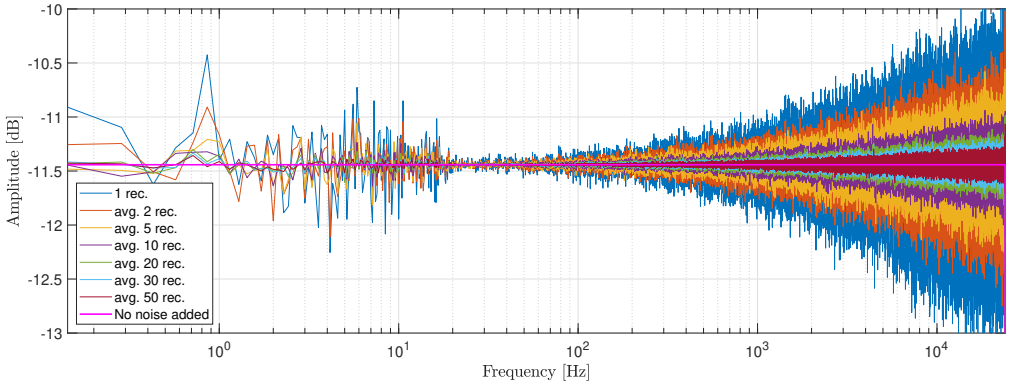


Figure D.2: Simulation showing the effect of averaging over more measurements in the Swept Sine Method. The simulation is made in MATLAB with a fixed seed `rng(1234)`.

D.2 Measuring Procedure

For readability, the plots and data used are shown for one microphone in a measurement setup where the dark zone are located in $[-0.25, 0.5]$ and the bright zone in $[0.25, 1.5]$, both relative to the loudspeaker array. The measurements of the impulse responses are done automatically. The input signal is a chirp from 20 Hz to 24 kHz (half of the sampling frequency 48 kHz). The chirp is then windowed with a shifted Hanning window with a length of 1.5 seconds. This is done in order to avoid playing more energy in the low frequencies than the loudspeakers can handle. The duration of the signal is 7 seconds with 1 second pre-delay, 5 seconds chirp and 1 seconds post-delay. The input signal can be seen in figure D.3.

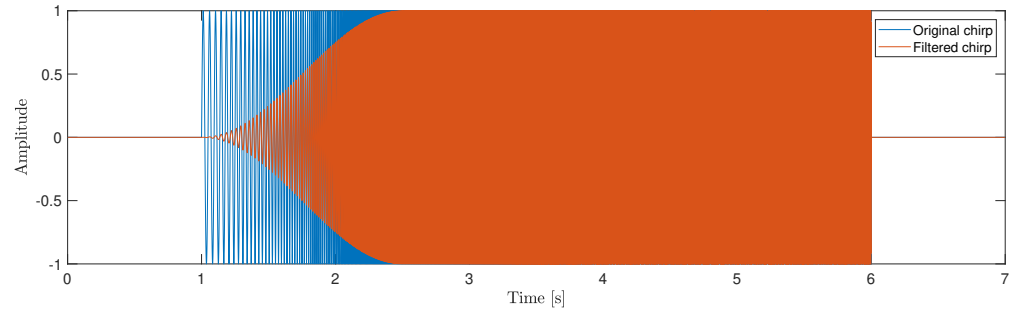


Figure D.3: Plot of the input signal in the Swept Sine Method.

The transfer functions from a source to all the control points are measured one at a time. The input signal are played back in the loudspeaker and simultaneously recorded by all the microphones in the control points. This is done five times for each loudspeaker.

D.3 Data Processing and Results

All the recordings are then made to impulse responses with the Swept Sine Method and saved in a cube containing: *number of control points* × *number of loudspeakers* × *length of the impulse*. The recorded signals are filtered with a second order high pass Butterworth filter with a cutoff frequency at 50 Hz, before making the impulse response. Due to a continuous low frequency noise from the microphones. This can be seen in figure D.4.

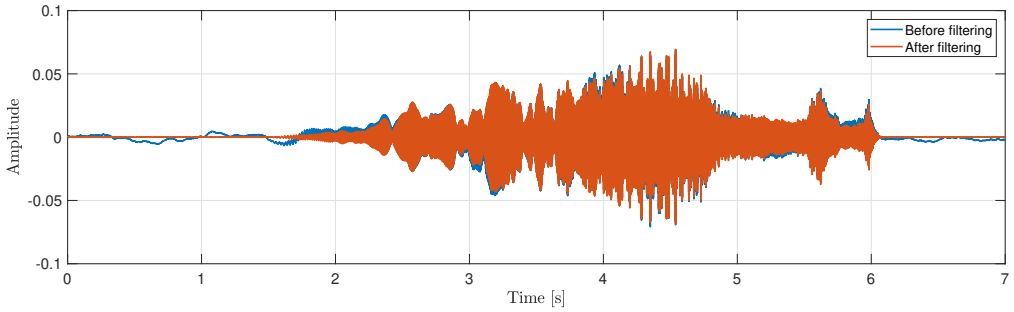


Figure D.4: Results of applying a second order Butterworth filter to the recorded signal.

An example of an impulse can be seen in figure D.5.

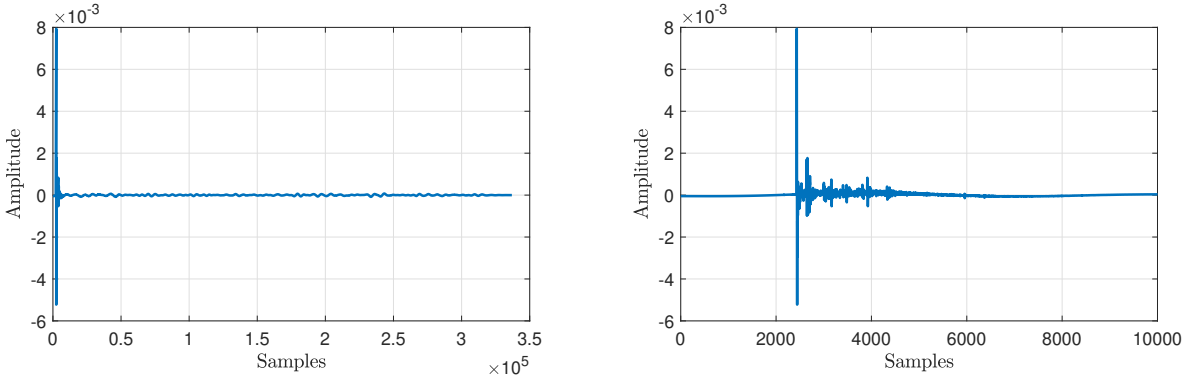


Figure D.5: Example of an impulse response created using the Swept Sine Method.

With the impulse responses for the room, the reverberation time can be found. However, reverberation time will only be calculated for the control points, meaning no spatial averaging of the room. The decay curve is made with Schroeder's backward integration, and the reverberation time can be found using T30, both described in [34, 35]. The results for one source/microphone combination can be seen in figure D.6, and the mean reverberation time for all the source/microphone combinations is approximately 0.18 seconds for frequencies 200 Hz to 5 kHz.

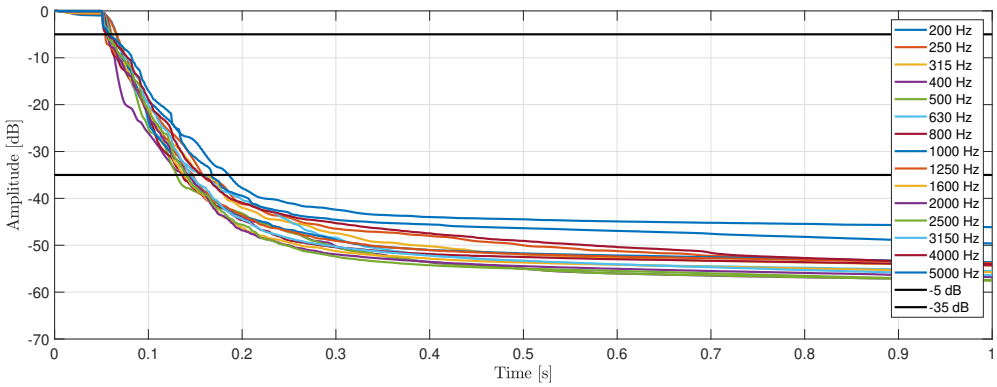


Figure D.6: Decay curves for octave band 200 Hz - 5 kHz created using Schroeder's backward integration. The two black lines indicates T30.

E | Simulated Contrast Shown as Amplitude Responses

The contrast plots from section 5.3, with the different ATF-models and different RSES-settings, shown as frequency plots in figures E.1, E.2, E.3 and E.4.

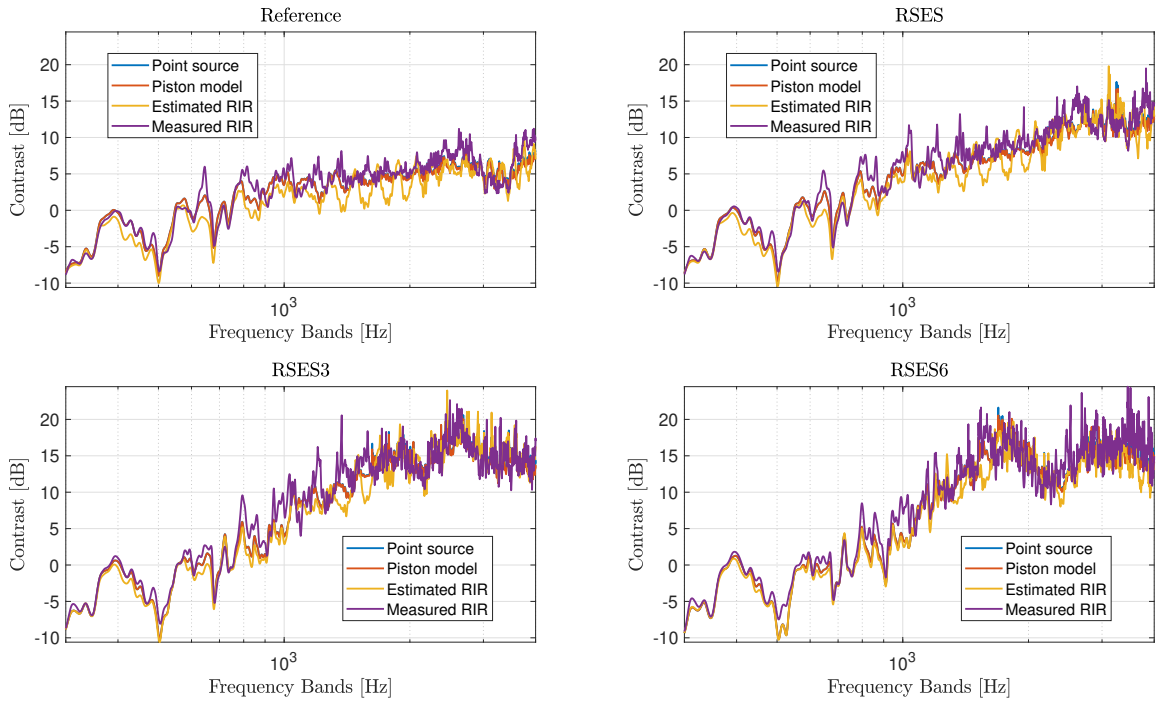


Figure E.1: Comparison of different models performance in order to achieved contrast when convolving the calculated filters with the measured impulse responses in the room. The setup is the one described in section 5.1, where the placement of the microphones are, relative to the loudspeaker array, $[0, 1.5]$ for the bright zone center and $[0, 0.5]$ for the dark zone center.

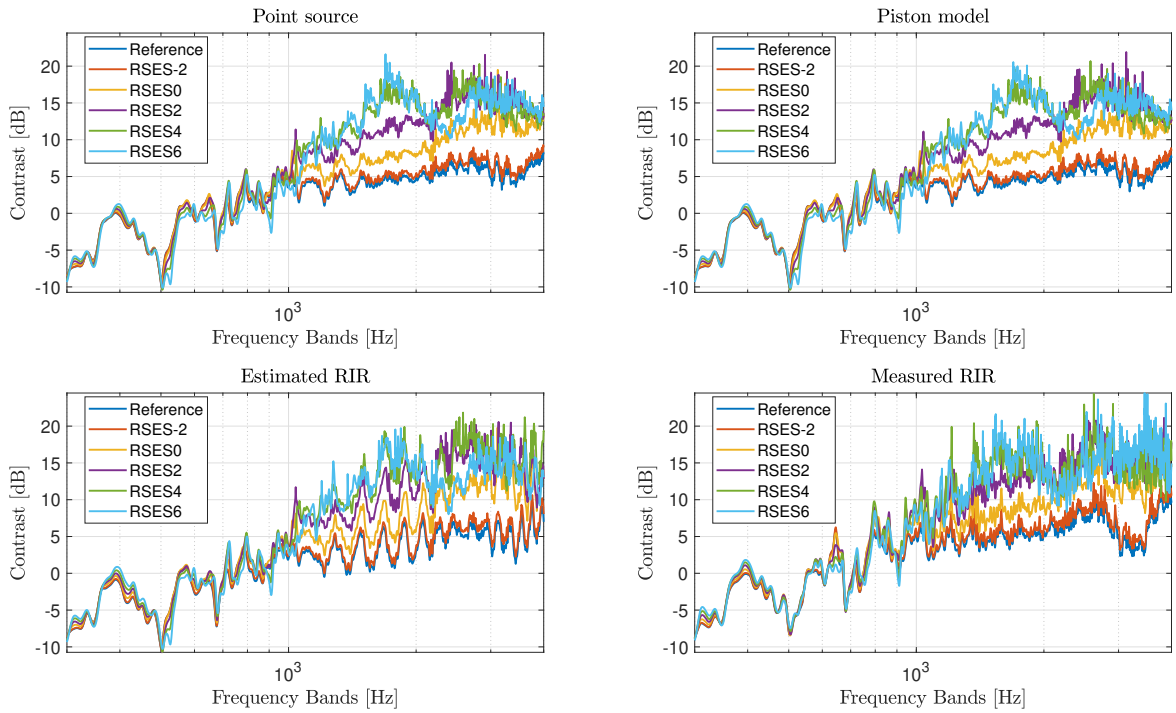


Figure E.2: Showing the effect of using the RSES-model with different settings to turn off loudspeakers. The placement of the microphones are, relative to the loudspeaker array, $[0, 1.5]$ for the bright zone center and $[0, 0.5]$ for the dark zone center.

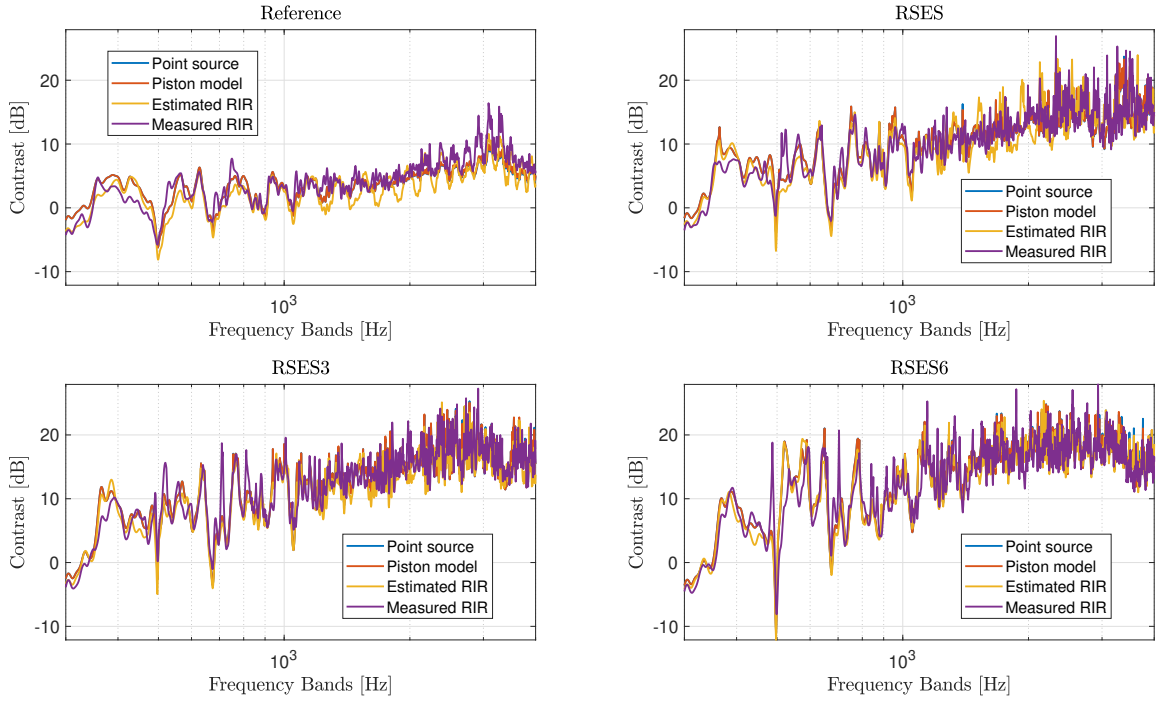


Figure E.3: Comparison of different models performance in order to achieved contrast when convolving the calculated filters with the measured impulse responses in the room. The setup is the one described in section 5.1, where the placement of the microphones are, relative to the loudspeaker array, $[0, 1.5]$ for the bright zone center and $[0, 0.5]$ for the dark zone center.

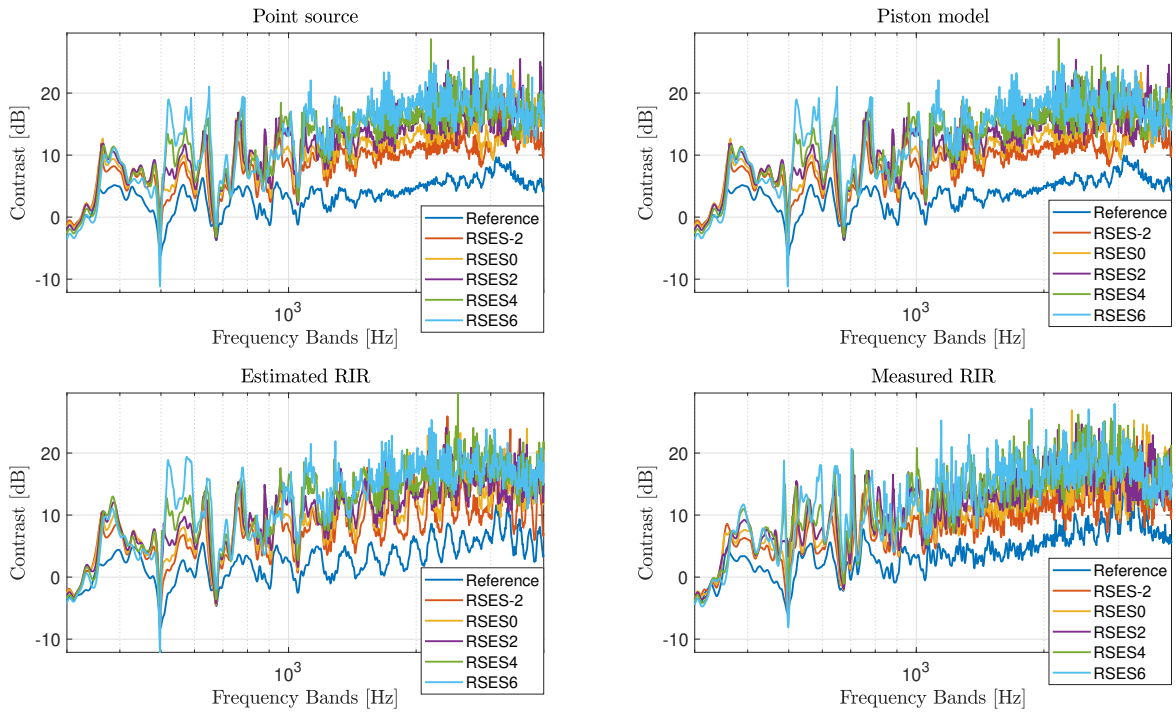


Figure E.4: Showing the effect of using the RSES-model with different settings to turn off loudspeakers. The placement of the microphones are, relative to the loudspeaker array, $[0, 1.5]$ for the bright zone center and $[0, 0.5]$ for the dark zone center.

F | Spectrograms of the Measurement Results

Spectrograms of the different measurements with different values of λ and different RSES-settings, shown in figures F.1, F.2, F.3 and F.4.

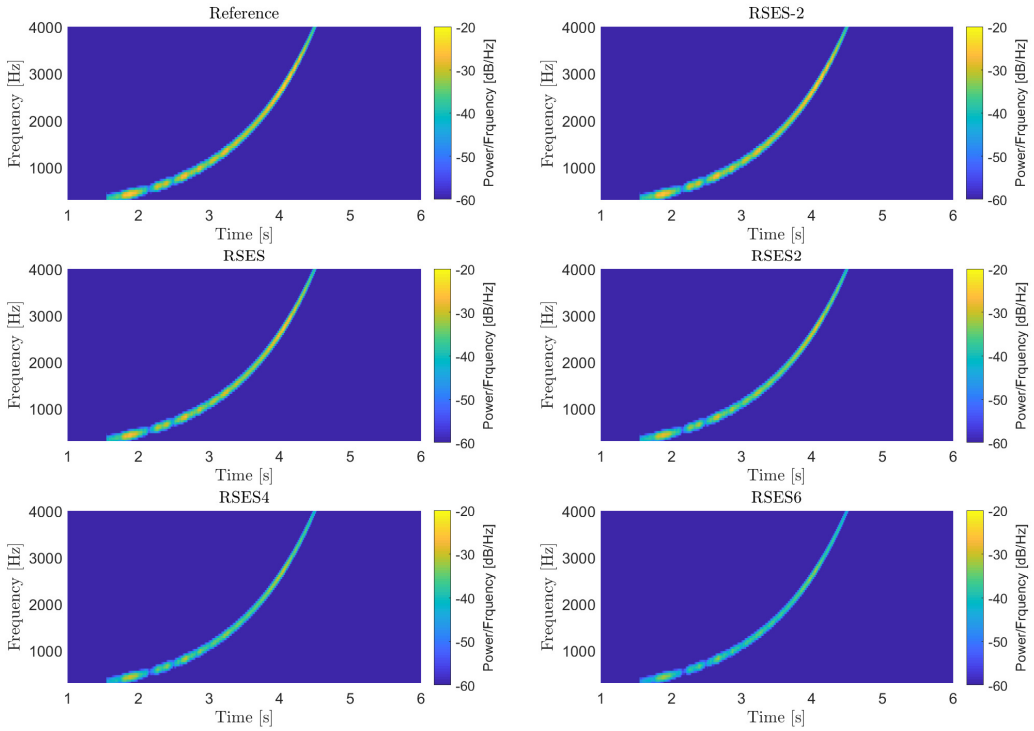


Figure F.1: Spetrogram for the measured results using **setup 1** with $\lambda = 0.1$, shown for different settings of the RSES-model.

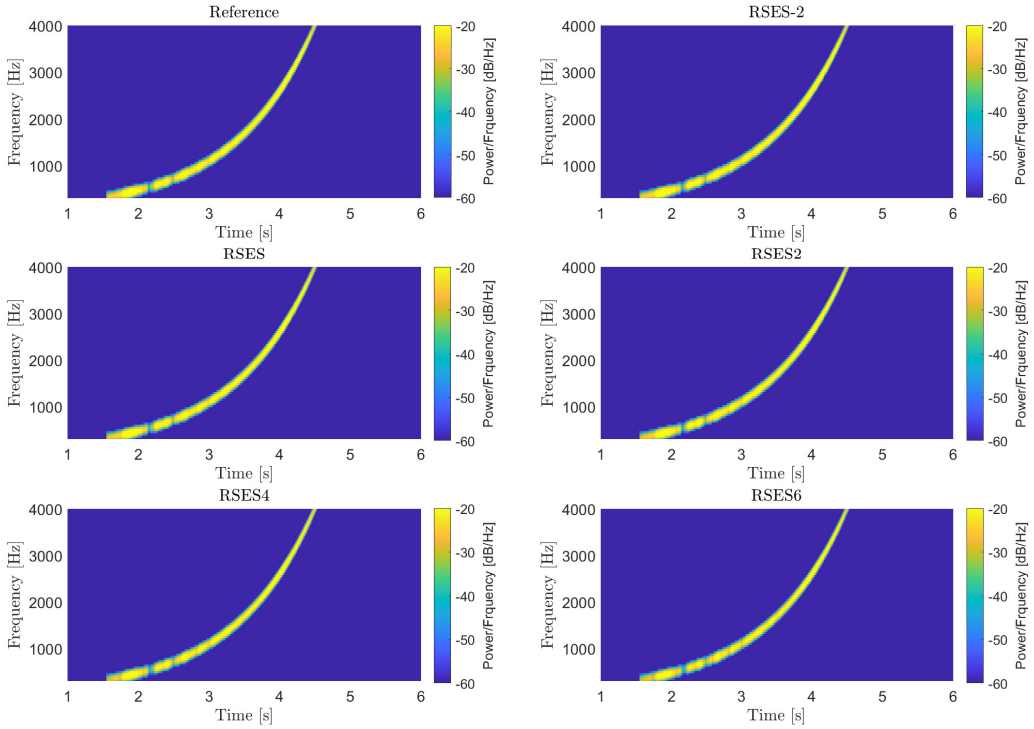


Figure F.2: Spetrogram for the measured results using **setup 1** with $\lambda = 0.01$, shown for different settings of the RSES-model.

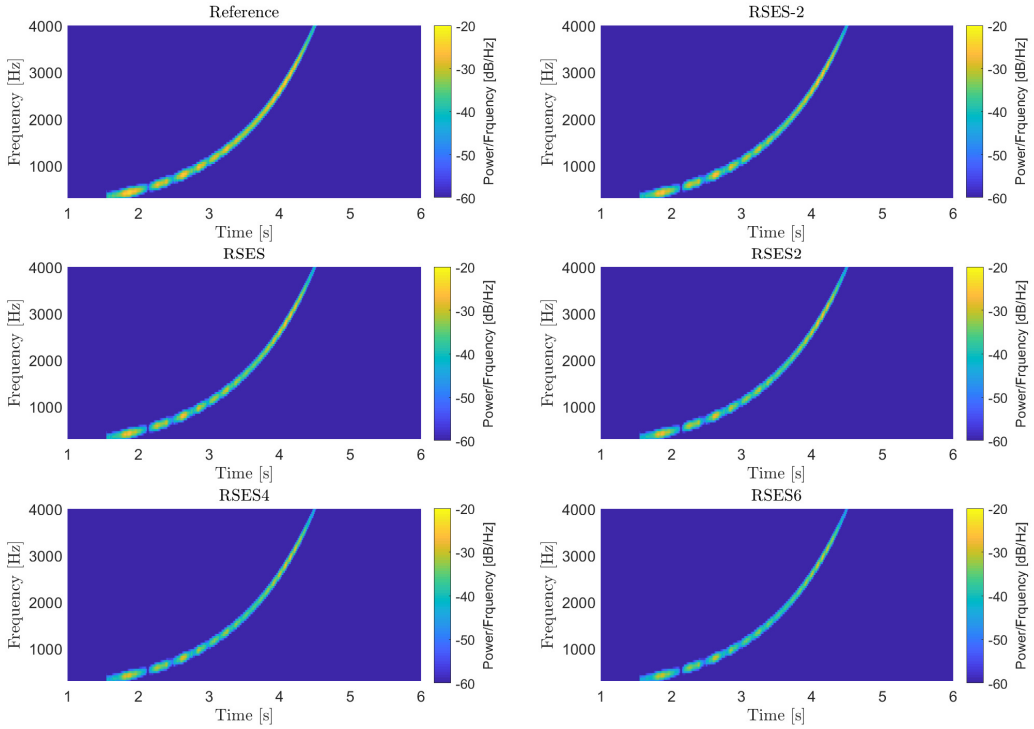


Figure F.3: Spetrogram for the measured results using **setup 2** with $\lambda = 0.1$, shown for different settings of the RSES-model.

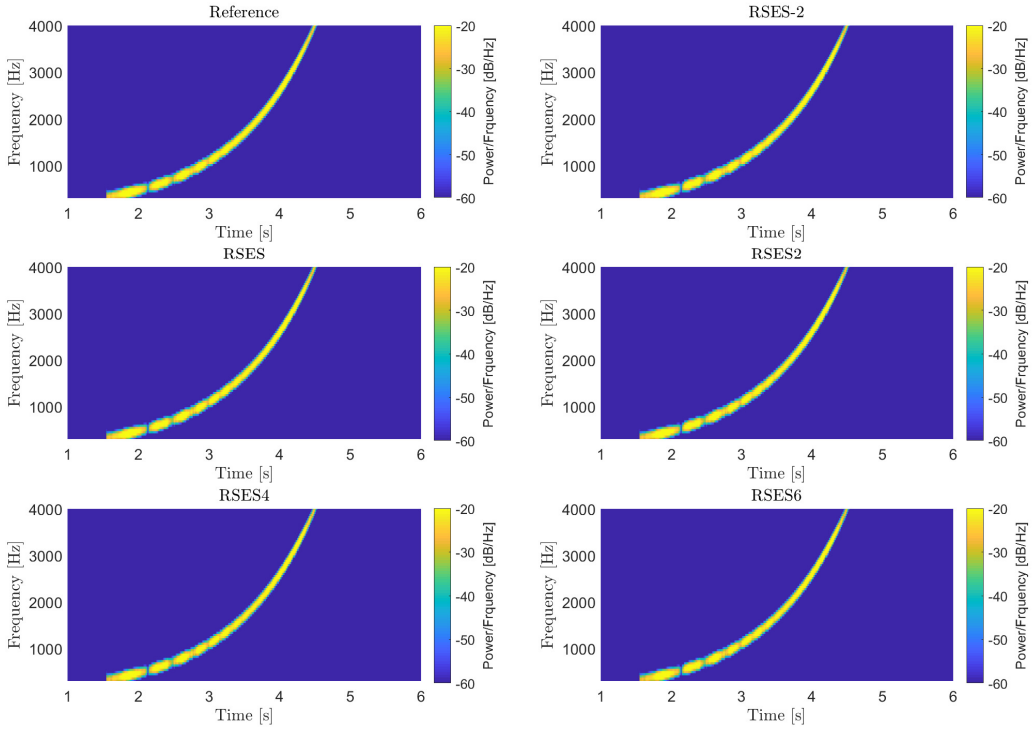


Figure F.4: Spetrogram for the measured results using **setup 2** with $\lambda = 0.01$, shown for different settings of the RSES-model.