Master's Thesis

Physical Structures for Signal Separation

Albin Nordsjö

Department of Electrical and Information Technology,
Faculty of Engineering, LTH, Lund University, 2016.

Physical Structures for Signal Separation

Albin Nordsjö
tfy11ano@student.lu.se

Department of Electrial And Information Technology
Lunds Tekniska Högskola

Advisors: Mikael Swartling, Nedelko Grbić

June 23, 2016

Abstract

This master's thesis studies the use of physical structures for signal source separation. A parabolic reflector was used to alter the directional properties of one microphone in a two-microphone array. A method to estimate the mixing channels from measured data is presented, as well as a method to find the optimal separating channels. Measurements of the directional properties and mixing channels in a two-source, two-sensor set-up were made using three reflectors of diameters 10, 5 and 2.5 cm. Speech and noise was mixed with the estimated channels and then un-mixed with the optimal separating channels. The performance of the separation was objectively evaluated, and the results were used to determine the impact of using reflective structures. The results showed that the 2.5 and 5 cm reflectors had very little effect for sound in the voice frequency range, but the 10 cm reflector improved the separation to a certain extent. Thus, there is a potential in using physical structures for signal separation purposes, however the frequency of the sound puts a limit on how small they can be made.

"To stay awake all night adds a day to your life," Stilgar said, accepting the tray with coffee as it was passed in the door.

Frank Herbert, *The Children of Dune*

Acknowledgements

My sincerest gratitude go to my advisors Dr. Mikael Swartling and Dr. Nedelko Grbić, for providing support, useful comments and suggestions that all contributed to the completion of this thesis. Special thanks to Dr. Grbić for suggesting this very interesting thesis proposal, which has been truly educational and rewarding to work with. Last but not least, I would like to thank my family for supporting and encouraging me throughout the years which I have studied.

Albin Nordsjö

Contents

List of Figures

List of Tables

List of Abbreviations

dB Decibel

DFT Discete Fourier Transform

FFT Fast Fourier Transform

FIR Finite Impulse Response

HATS Head And Torso Simulator

MEMS Micro-Electro-Mechanical System

MOS Mean Objective Score

PESQ Perceptual Evaluation of Speech Quality

PSD Power Spectral Density

SNR Signal-to-Noise Ratio

WOSA Weighted Overlap Segmented Averaging

Chapter 1
Introduction

1.1 Background

Algorithm based methods to create an adaptive directional microphone array has previously been proposed, for example in [1]. Here, an array of two microphones is used to generate a minimum in a certain direction. Direction-of-arrival estimation using a single microphone and a parabolic reflector has been studied in [2]. The directional properties of the human ear and how the brain interprets this information to determine a sound source direction is also a current topic of research [3] [4] [5]. Most approaches are either algorithmic or physical, and the combination of an array with multiple microphones and physical structures is relatively unexplored.

The idea of using a parabolic reflector to obtain a directional microphone dates back to the days before the radar was invented. Tests were made to listen for incoming aircraft using big reflectors, however when the radar was invented it replaced all sonic aircraft detection systems. Parabolic reflectors were used in early sound recording for movie and television, but was later replaced by boom microphones. Today, parabolic reflector microphones are mainly used to record birdsong and wildlife and to capture sounds on the playing field of various sports.

For this thesis two microphones will be placed in an array, with a parabolic reflector between them. The first microphone will be placed in the focal point of the parabolic reflector, and the second one will be placed behind the reflector. A sketch of the array can be seen in Figure 1.1. The channel estimation method will also be used on a pair of compact communication earpieces made by the company EarIn.

1.2 Objectives

The objective of this thesis is to investigate how a physical structure can be used in a microphone array to change its directional properties, how this

affects the mixing of the recorded signals and how that in turn affects the separation of the signals.

1.3 Thesis outline

This thesis report is divided into six chapters. In the first chapter the background and objective of the thesis are presented. In the second chapter the theoretical background is provided. First different microphones and their directional properties are reviewed, then the mixing model and the method used to identify the mixing channels are presented. The parabolic reflector is then described, followed by a method to calculate the inverse of the mixing channels. Finally there is an overview of the Preceptive Evaluation of Speech Quality (PESQ), which is a way to objectively evaluate the quality of a speech recording. In the third chapter the mixing channels and their determinants are calculated for microphone arrays with different parabolic reflectors. In the fourth chapter the measurement equipment and methodology is presented, followed by the results of the measurements that were made. In the fifth chapter the outcome of the measurements are discussed, and conclusions regarding whether or not a parabolic reflector can be used to improve source separation is drawn. In the sixth and final chapter some suggestions on further research that could be made based on the outcome of this thesis are provided.

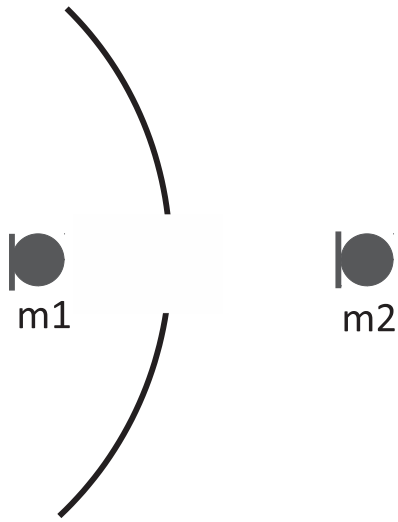

Figure 1.1: A sketch of the array, with two microphones and a parabolic reflector.

Chapter 2
Theory

2.1 Microphone types

2.1.1 Condenser microphones

Most modern measurement microphones are pressure-sensing condenser microphones. They operate by the variation of the distance between a movable and stationary plate which are electrically conducting and either charged by an applied voltage or a permanent charge in one of the plates, thus forming a capacitor. One of the plates is exposed to the air and the other is fixed inside the microphone body. The exposed plate consists of a thin stretched diaphragm which moves as the air pressure changes, thus altering the capacitance. The change in capacitance is converted to a proportional voltage by the microphone's electronics. The condenser microphone thus requires a voltage supply to operate [6]. The advantage of the condenser microphone is that the moving element has a low mass, making it sensitive to high frequency air-pressure changes [7].

2.1.2 Dynamic microphones

Dynamic microphones make use of the principle that a changing magnetic field induces current into a conductor placed in that field. The change could be either the strength of the field or the conductors position within the field. For moving-coil microphones, the conductor is a thin coil of wire placed in a fixed magnetic field and attached to a diaphragm in contact with the air. As the air pressure varies, the diaphragm and the coil moves, inducing a current in the wire. The voltage output is proportional to the velocity of the diaphragm rather than the position as for the condenser microphone. The moving coil microphone is simple in principle, but implementing it in reality requires careful considerations to achieve a good frequency response. The mass of the coil gives it a certain inertia, which affects its sensitivity, especially to high frequency changes in air velocity. The diaphragm will

also respond differently depending on the direction of the incoming sound [7].

Another type of dynamic microphone is the ribbon microphone, in which a thin strip of corrugated metal is placed in a magnetic field. The ribbon moves in response to the air movement and the output voltage level is proportional to the air velocity [8] [9].

2.1.3 Other types of microphones

Some other types of dynamic microphones are the carbon microphone, the piezoelectric microphone, the laser microphone and the MEMS microphone. The carbon microphone consists of a capsule containing carbon granulates pressed between two metal plates. Changes in the air pressure deforms the granulates causing the contact area between adjacent granules to change, which in turn causes the electrical resistance of the capsule of carbon granulates to change [9]. Carbon microphones were widely used in telephones from 1890 until the 1980's. The carbon microphones can produce high level audio signals from very low voltage input. Today they are mainly used in safety critical applications such as mining and chemical manufacturing, where sparks produced by higher voltages could cause accidents. They are also used in emergency military communication channels since they are insensitive to voltage peaks from lightning strikes and the electromagnetic pulse generated by a nuclear explosion [10].

In a piezoelectric microphone voltage is generated by the deformation of a crystal with piezoelectrical properties [9]. They are generally used as contact microphones for example on acoustical musical instruments or in high pressure environments underwater [11].

The laser microphone works by illuminating a diaphragm with a laser beam, detecting the changes in the reflected light using an interferometer. It can be used for surveillance, since sound can be detected over long distances, but also for detection of molecules and seismic studies. [12]

The micro-electro-mechanical, or MEMS microphones for short is a relatively new type of microphone. It consists of a diaphragm etched directly onto a silicon wafer using MEMS manufacturing techniques. Most MEMS microphones record sound using the same principle as a condenser microphone. The MEMS microphone is a hot topic of research and can be found in a wide number of applications such as mobile phones, computers, and wearable computers, bluetooth headsets, hearing aids, consumer electronics and in automotive voice control [13].

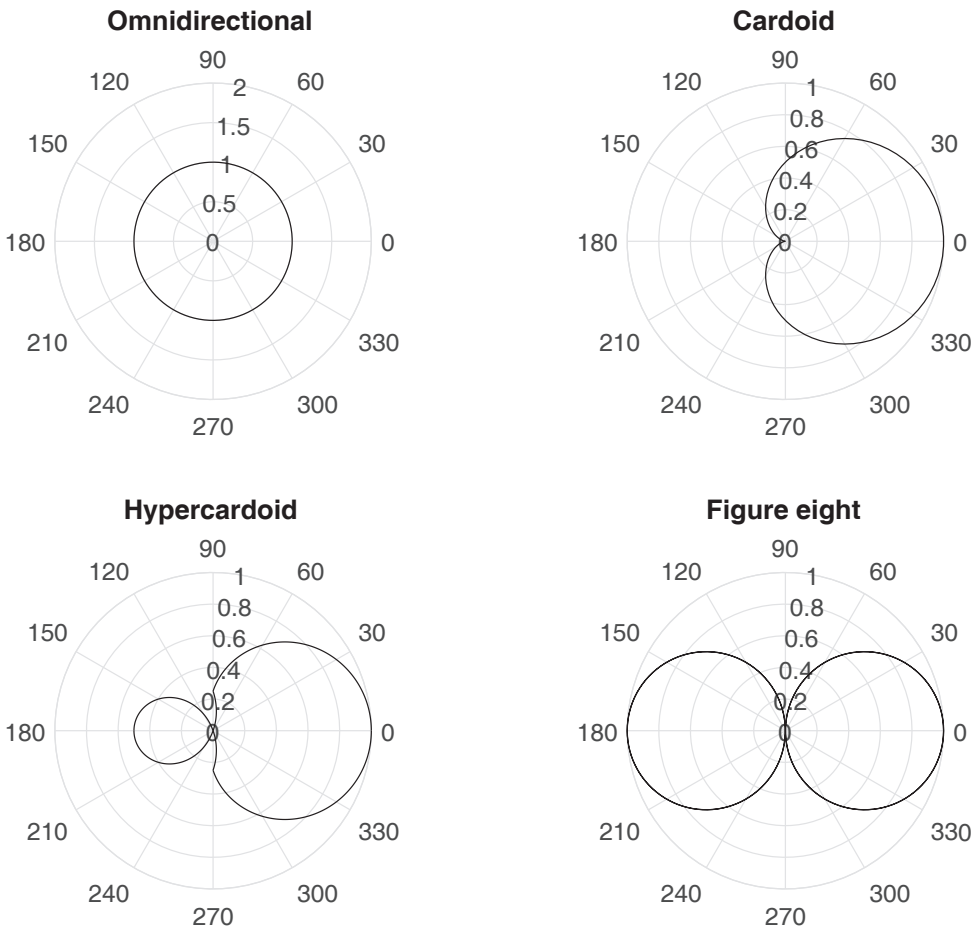

Figure 2.1: A few different polar patterns for microphones typically used for vocal and musical instrument recording.

2.1.4 Directional properties of microphones

Depending on both the diaphragm construction and the housing of the microphone it can be either equally sensitive to sounds coming from any direction or more sensitive to sounds coming from a certain direction. If the microphone records absolute pressure it is always omnidirectional, but if it records the pressure gradient from front to back it has directionality [7]. Condenser and moving coil microphones typically have a cardoid or hypercardoid polar pattern and the ribbon microphone has a figure eight sensitivity pattern. The different polar patterns are illustrated in Figure 2.1.

A few different structures to alter the directionality of a microphone has been developed. The so called *machinegun* and *shotgun* microphones uses

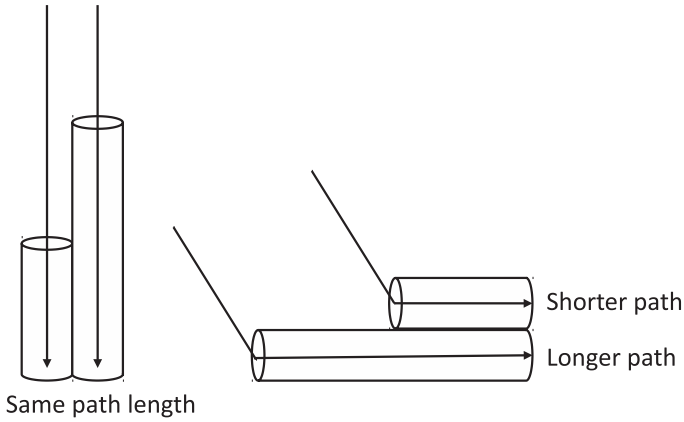

Figure 2.2: The principle for the machine gun and shotgun microphones.

interference tubes to change the directional property of the microphone, as illustrated in Figure 2.2. The machinegun microphone is constructed by placing a microphone at the end of a number of tubes of various length. The shotgun microphone is a development of the same concept. It uses a single tube with a number of holes or slits along its side [14], see Figure 2.3. The purpose of this is to generate destructive interference for sounds coming in from the side of the microphone. Sound coming from the front will go straight through the tube. Sound coming from the sides will enter the tube both through its end and through the holes on its side. This means that the sound will have different path lengths to the microphone and thus it will cause interference.

The directional properties of a microphone can also be altered by putting a structure of a reflecting material around the microphone. For example a parabolic reflector, which will be further examined in the thesis. The parabolic reflector reflects the sound coming in from the front of the reflector onto a single point, the focal point of the reflector.

2.2 Convoluted mixtures

Blind source separation is the process of recovering unobserved independent sources from observed mixtures of these sources. The simplest case is the instantaneous mixing model, where the sources are linearly superimposed by the mixing channels [15]. This is a well studied problem, and many algorithms that can solve it has been proposed.

The instantaneous mixing model is described as

$$\mathbf{x}(t) = A\mathbf{s}(t) \tag{2.1}$$

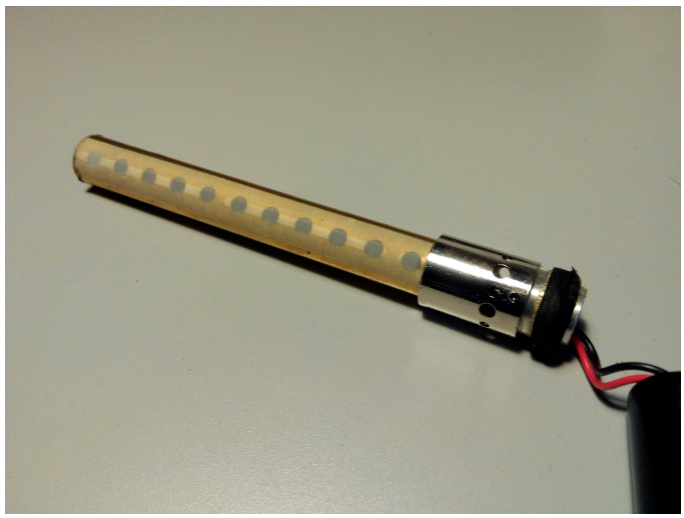

Figure 2.3: The interference tube in a shotgun microphone [16].

where A is the mixing matrix, $\mathbf{s}(t) = [s_1(t)\ s_2(t)\ ...\ s_N(t)]^T$ are the source signals and $\mathbf{x}(t) = [x_1(t)\ x_2(t)\ ...\ x_N(t)]^T$ are the observed signals. The objective is to estimate the inverse of the mixing matrix A^{-1} based on the observations $\mathbf{x}(t)$ so that estimates of the original sources $\mathbf{s}(t)$ can be formed.

An extension to the instantaneous mixing model is the convoluted mixing model. In this model the observed discrete time signal $x_j(n)$ is generated from the unknown source signals $\mathbf{s}(n)$ by the convolution model

$$x_j(n) = \sum_{i=1}^{N} \sum_{k=-\infty}^{\infty} a_{k,i} s_i(n-k) \tag{2.2}$$

thus a number of delayed versions of the source signals are mixed together. Here, both the source signals $\mathbf{s}(n)$ and the convolution coefficients $a_{k,i}$ are unknown. We want to estimate the source signals $\mathbf{s}(n)$ by using the observations $\mathbf{x}(n)$ and find a deconvolution filter such that

$$y_j(n) = \sum_{i=1}^{N} \sum_{k=-\infty}^{\infty} w_{k,i} x_i(n-k) \tag{2.3}$$

is a good estimate of the source signal $s_j(n)$ at each time instant. This achieved by choosing the coefficients $w_{k,i}$ of the deconvolution filter suitably [17].

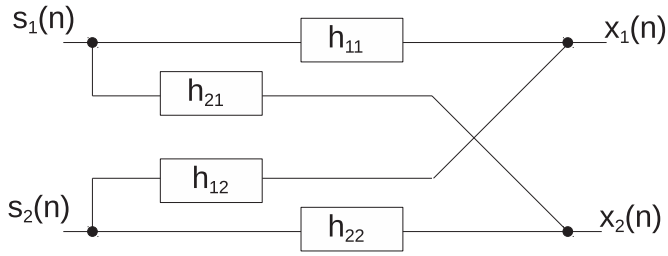

Figure 2.4: A block diagram describing the mixing channels.

2.2.1 The Mixing Matrix

In an array set-up without a reflector, all the components of the mixing matrix will typically be of similar magnitude only differed by phase. Adding a reflector such that it reflects only one of the sources into one of the microphones will add an amplified and additionally delayed entry to one of the components of the mixing matrix.

If the speech source and the noise source is close to one and other in strength and possibly even similar in structure, the source separation can be very hard. An example is the classic cocktail party problem of identifying the voice of single speaker from a convoluted mix of a number of people talking in a room at the same time.

Instantaneous mixtures can generally be solved, but in a real environment where echoes generate convoluted mixtures, many of the modern algorithms fail or perform poorly. It is possible that by using a reflective structure, the mixing matrix could be changed in such a way that there is additional diversity between the speech source and the noise sources, which might improve the source separation.

2.2.2 Reflective structure

A system with two sources and two sensors as shown in Figure 2.4, can be described using the convoluted mixture model as

$$\begin{pmatrix} \mathbf{x_1}(n) \\ \mathbf{x_2}(n) \end{pmatrix} = \begin{pmatrix} \mathbf{h_{11}}(n) & \mathbf{h_{12}}(n) \\ \mathbf{h_{21}}(n) & \mathbf{h_{22}}(n) \end{pmatrix} * \begin{pmatrix} \mathbf{s_1}(n) \\ \mathbf{s_2}(n) \end{pmatrix} \tag{2.4}$$

Where $\mathbf{h_{11}}(n)$, $\mathbf{h_{12}}(n)$, $\mathbf{h_{21}}(n)$ and $\mathbf{h_{22}}(n)$ are arrays describing the filters caused by the propagation paths for signals picked up by the different sensors. In addition to this, when a structure is added, $\mathbf{h_{11}}(n)$ will also model the reflection and amplification caused by this structure.

The filter $\mathbf{h_{11}}(n)$ describes propagation path to the first microphone, and will be a combination of the arrival time of the direct wave, and the slightly longer arrival time of the reflected wave. The filters $\mathbf{h_{12}}(n)$, $\mathbf{h_{21}}(n)$ and $\mathbf{h_{22}}(n)$ will ideally only have entries representing direct waves.

To model a situation with one speech source and one noise source, the sources are placed in a two-dimensional plane, and the time it takes the sound to travel from the speech source and from the noise source to both of the two microphones are calculated. The additional delay and amplification of the reflected sound is also calculated. These travel times are then converted to numbers of samples by multiplying with the sample rate. Generally this will not yield an integer number of samples.

This is modelled by creating the convolution coefficients a_k using $\mathrm{sinc}(k-\delta)$, where δ is the delay in fractional samples. This replicates what happens when a signal that is delayed a fractional number of samples is recorded by the microphone and A/D-converted.

$$\mathrm{sinc}(x) = \begin{cases} \dfrac{\sin(\pi x)}{\pi x} & \text{if } x \neq 0 \\ 1 & \text{if } x \equiv 0 \end{cases}$$

2.2.3 The determinant of the mixing matrix

The inverse of a square matrix A is given by [18]

$$A^{-1} = \frac{1}{\det A}\,\mathrm{adj}(A) \tag{2.5}$$

To determine the possibilities of finding an inverse to the mixing matrix, the determinant of the mixing matrix will be studied. A matrix with a determinant equal to zero is not invertible, and if the determinant is very small the inverse will be very big, and thus sensitive to errors. The smaller the determinant the more ill-conditioned the problem becomes, and thus it is harder to solve accurately [19]. The idea with using a reflector is to change the mixing matrix in such a way that the absolute value of the determinant increases, thus making the problem better conditioned. The mixing matrix in the frequency domain is obtained by taking the discrete Fourier transform of the impulse responses. $H_{kj}(f_k)$ is the discrete Fourier transform of $h_{kj}(n)$ and the frequency bins $f_k = k/N \cdot F_s$ where N is the number of DFT points, F_s the sample frequency and $k = [0,\ 1,\ ...,\ N-1]$. The mixing matrix in the frequency domain is

$$H(f_k) = \begin{pmatrix} H_{11}(f_k) & H_{12}(f_k) \\ H_{21}(f_k) & H_{22}(f_k) \end{pmatrix} \tag{2.6}$$

The determinant will be calculated for each frequency independently as

$$\det(H(f_k)) = \begin{vmatrix} H_{11}(f_k) & H_{12}(f_k) \\ H_{21}(f_k) & H_{22}(f_k) \end{vmatrix} \tag{2.7}$$

$$\Longleftrightarrow$$

$$\det(H(f_k)) = H_{11}(f_k)H_{22}(f_k) - H_{12}(f_k)H_{21}(f_k) \tag{2.8}$$

The smaller the difference between $H_{11}(f_k)H_{22}(f_k)$ and $H_{12}(f_k)H_{21}(f_k)$, the smaller the absolute value of the determinant will be. The goal is for the reflector to only affect $H_{11}(f_k)$, keeping $H_{12}(f_k)$, $H_{21}(f_k)$ and $H_{22}(f_k)$ unchanged. If $H_{11}(f_k)$ can be increased by using a reflector, the determinant will also increase, and the problem will be better conditioned. In the same way, if $H_{11}(f_k)$ is decreased so that the determinant is less than zero, the problem will also be better conditioned. The imaginary parts of the frequency response functions are linked to the phase difference between the microphones. By changing the distance between the microphones the phase difference changes. There will be an addition to the imaginary part in $H_{11}(f)$ representing the phase difference caused by the additional distance the reflected wave has to travel.

2.3 Channel identification

2.3.1 Spectral density

The spectral density of a data set can be calculated using the Fourier transform. For two stationary random processes $x(t)$ and $y(t)$, the short term Fourier transforms over the k:th record of data of length T can be calculated as

$$X_k(f,T) = \int_0^T x_k(t)e^{-j2\pi ft}\, dt \tag{2.9}$$

The spectra $Y_k(f,T)$ is calculated the same way, replacing $x_k(t)$ with $y_k(t)$. The cross-spectral density function between the two random processes is defined as

$$S_{xy} = \lim_{T\to\infty} \frac{1}{T} E[X_k^*(f,T)Y_k(f,T)], \tag{2.10}$$

where $X_k^*(f,T)$ denotes the complex conjugate of $X_k(f,T)$. The one-sided cross spectra is given by

$$G_{xy}(f) = \lim_{T\to\infty} \frac{2}{T} E[X_k^*(f,T)Y_k(f,T)], \quad f > 0 \tag{2.11}$$

The auto spectra $G_{xx}(f)$ is calculated in the same manner, with $X_k(f,T) = Y_k(f,T)$. A constant-parameter linear system with a weighting function

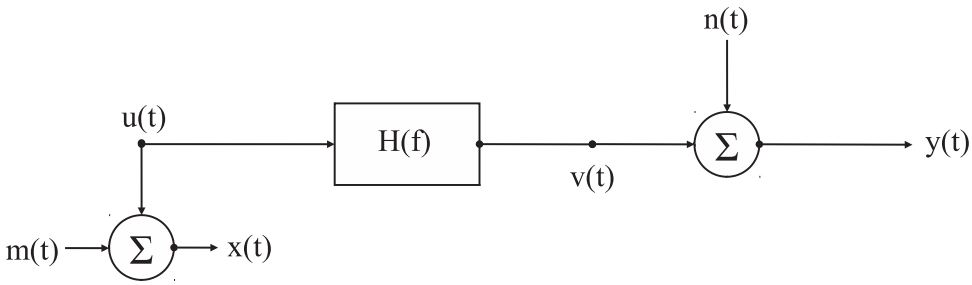

Figure 2.5: Single-input/single-output system with input and output noise.

$h(\tau)$ and frequency response function $H(f)$, subjected to the well-defined single input $x(t)$, produces a well-defined output $y(t)$. This system can be described by the convolution integral [20]

$$y(t) = \int_0^\infty h(\tau)x(t-\tau)\ d\tau \tag{2.12}$$

The one-sided spectral density functions are related as

$$G_{yy}(f) = |H(f)|^2 G_{xx}(f) \tag{2.13}$$

$$G_{xy}(f) = H(f)G_{xx}(f) \tag{2.14}$$

2.3.2 The H_1 estimator

When noise is present at both the input and the output of the system, as shown in Figure 2.5, the measured input and output will be

$$x(t) = u(t) + m(t) \tag{2.15}$$

$$y(t) = v(t) + n(t) \tag{2.16}$$

where $u(t)$ and $v(t)$ are the true signals, and $m(t)$ and $n(t)$ are noise terms. The input and output noise, $m(t)$ and $n(t)$, are assumed to be uncorrelated both to each other, and to the signals $u(t)$ and $v(t)$, i.e.

$$G_{um}(f) = G_{vn}(f) = G_{mn}(f) = 0 \tag{2.17}$$

Since $u(t)$ and $v(t)$ are polluted by noise, we cannot directly obtain their spectral density functions, the measurable spectral density functions, $G_{xx}(f)$, $G_{yy}(f)$ and $G_{xy}(f)$ are

$$G_{xx}(f) = G_{uu}(f) + G_{mm}(f) \geq G_{uu}(f) \tag{2.18}$$

$$G_{yy}(f) = G_{vv}(f) + G_{nn}(f) \geq G_{vv}(f) \tag{2.19}$$

$$G_{xy}(f) = G_{uv}(f) \tag{2.20}$$

since $G_{mm}(f) \geq 0$ and $G_{nn}(f) \geq 0$ for all f.

Assuming that the input noise $m(t)$ is negligible, and the output noise is uncorrelated, an unbiased estimate of the system frequency response, $H_1(f)$ is given by [20]

$$H_1(f) = \frac{G_{xy}(f)}{G_{xx}(f)} \tag{2.21}$$

2.3.3 Optimality of the H_1 estimator

Consider the system in Figure 2.5, assuming that the input noise $m(t) = 0$. Let $H(f)$ be *any* linear frequency response function. For long records of time T the equation describing Figure 2.5 is

$$Y(f,T) = H(f)X(f,T) + N(f,T) \tag{2.22}$$

where $X(f,T)$, $Y(f,T)$ and $N(f,T)$ are the Fourier transforms of $x(t)$, $y(t)$ and $n(t)$ respectively. It follows that

$$N(f,T) = Y(f,T) - H(f)X(f,T) \tag{2.23}$$

and

$$\begin{aligned} |N(f,T)^2| = |Y(f,T)|^2 - H(f)X(f,T)Y^*(f,T) \\ -H^*(f)X^*(f,T)Y(f,T) + H(f)H^*(f)|X(f,T)|^2 \end{aligned} \tag{2.24}$$

Taking the expectation of (2.24), using (2.11), multiplying with $2/T$, and letting T increase to infinity yields

$$\begin{aligned} G_{nn}(f) = G_{yy}(f) - H(f)G_{xy}(f) - H^*(f)G_{xy}(f) \\ + H(f)H^*(f)G_{xx}(f) \end{aligned} \tag{2.25}$$

The optimum choice of $H(f)$ is the one that minimizes $G_{nn}(f)$ over all possible $H(f)$. The dependence on f will be omitted to simplify the derivation.

Let the complex numbers be expressed in terms of their real and imaginary parts:

$$\begin{array}{ll} H = H_R - jH_I & H^* = H_R + jH_I \\ G_{xy} = G_R - jG_i & G_{yx} = G_R + jG_I \end{array} \tag{2.26}$$

This yields

$$G_{nn} = G_{yy} - (H_R - jH_I)G_{yx} - (H_R + jH_I)G_{xy} + (H_R^2 + H_I^2)G_{xx} \tag{2.27}$$

Taking the partial derivatives with respect to H_R and H_I and setting the equal to zero gives

$$\begin{aligned} \frac{\delta G_{nn}}{\delta H_R} &= -G_{yx} - G_{xy} + 2H_R G_{xx} = 0 \\ \frac{\delta G_{nn}}{\delta H_I} &= jG_{yx} - jG_{xy} + 2H_I G_{xx} = 0 \end{aligned} \tag{2.28}$$

which leads to

$$\begin{aligned} H_R &= \frac{G_{xy} + G_{yx}}{2G_{xx}} = \frac{G_R}{G_{xx}} \\ H_I &= \frac{j(G_{xy} - G_{yx})}{2G_{xx}} = \frac{G_I}{G_{xx}} \end{aligned} \tag{2.29}$$

which gives the optimal solution, which conforms to equation (2.21) [20],

$$H(f) = H_R(f) - jH_I(f) = \frac{G_R(f) - jG_I(f)}{G_{xx}(f)} = \frac{G_{xy}(f)}{G_{xx}(f)} \tag{2.30}$$

2.3.4 Coherence function

The coherence function is defined as [20]

$$\gamma_{xy}^2(f) = \frac{|G_{xy}(f)|^2}{G_{xx}(f)G_{yy}(f)}\ , \quad 0 \leq \gamma_{xy}^2(f) \leq 1 \tag{2.31}$$

For a linear time invariant system without input and output noise, the coherence function will be unity for all frequencies. As input and/or output noise increases, the coherence function will decrease. By studying the coherence function errors in the measurement process can be detected. The existence of bias errors in the estimates of the frequency response functions due to input noise, inadequate spectral resolution and non-linear effects will produce indicative anomalies in the coherence function. According to [20] some guidelines are:

1. If $\hat{\gamma}_{xy}^2(f)$ falls over a frequency range where $|\hat{H}(f)|$ is not near a minimum value, but $\hat{G}_{xx}(f)$ is near a minimum value, then measurement noise at the input should be suspected.

2. If $\hat{\gamma}_{xy}^2(f)$ notches sharply at a frequency where $|\hat{H}(f)|$ displays sharp peak or notch, then inadequate spectral resolution in the analysis is the most likely cause, although non-linearities can produce similar results at peaks in $|\hat{H}(f)|$.

3. To distinguish between resolution problems and non-linearities, the analysis should be repeated with an improved spectral resolution. In increased value of $\hat{\gamma}_{xy}^2(f)$ will confirm a resolution problem. Otherwise, non-linearities should be investigated.

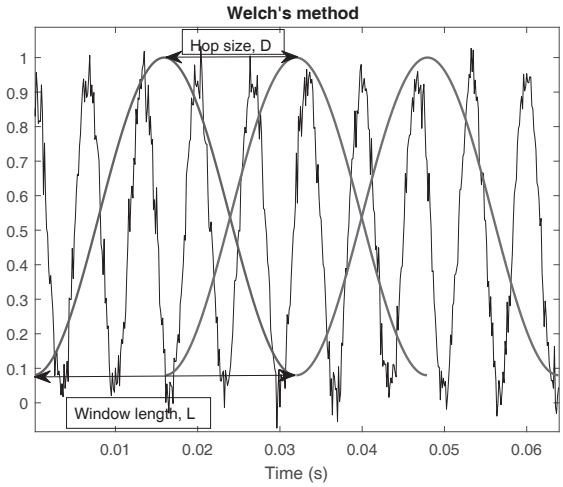

Figure 2.6: Applying the Welch window method on a noisy sine wave.

2.3.5 Welch window method for spectrum estimation

The spectral density functions $G_{xx}(f)$, $G_{yy}(f)$ and $G_{xy}(f)$ will be calculated using Weighted Overlap Segmented Averaging, WOSA as proposed by Peter D. Welch in 1967 [21]. It is also known simply as Welch's window method.

The principle of WOSA is shown in Figure 2.6. The signal is split up into N overlapping segments of length L, separated by the hop size D. The overlapping segments are then windowed in the time domain. In this implementation a Hamming window will be used. After segmenting and windowing the data, the Fourier transform of each windowed segment is calculated, and the cross correlation function, $g_{xy}(f,k) = X_k^*(f)Y_k(f)$, or in the case of auto correlation $g_{xx}(f,k) = |X_k(f,T)|^2$, is calculated. The average correlation functions are then calculated

$$G(f) = \frac{1}{N}\sum_{k=1}^{N} g(f,k) \tag{2.32}$$

By averaging several uncorrelated periodograms instead of transforming the whole data set in one go, the variance the spectral estimation is reduced, in exchange of reducing the frequency resolution. Welch's method reduces the impact of noise caused by imperfect and finite data [22].

2.3.6 Practical considerations on measurements and estimates.

The input will either be generated in Matlab or consist of known speech recordings. This signal will then be played back by a speaker placed inside the anechoic sound lab at LTH. There are a number of sources that could add noise to the input signal before it is played back by the speakers, but this noise is likely to be so small that it is negligible. Noise could be induced into the PC sound card due to the electromagnetically polluted environment inside the PC. By using an external sound card these disturbances can be reduced.

It is also likely that some noise is caused by the speaker due to the frequency response of the speaker not being entirely flat, and maybe some non-linear behaviour of the speaker causing for example some harmonic distortion. To cover the frequency range from 20 Hz to 20 kHz multiple speakers of different types are usually required, at least a bass, a mid range and a treble speaker. These experiments will be performed using mid range speakers, so accuracy for low and high frequencies can not be expected. One of the speakers is mounted in the Brüel & Kjaer HATS model. Since the speaker is placed inside of a cavity in the dummy head, it is very likely that some filtering of the signal will occur. By placing a microphone in front of the mouth of the HATS and using the sound from this microphone as the input signal when estimating the channels of the array, the filtering in the HATS is bypassed.

Another error source could be that some other unmeasured input contributes to the output e.g. some kind of background noise, however since the experiments will be preformed in a sound isolated anechoic room, it is unlikely that a source other than the one played back by the speakers should be present.

It is likely that there is more noise present in the output compared to the input. The main noise source in the output would probably be the recording microphones. The microphones used are very small and quite cheap, so a compromise between accuracy, size and prize must have been made by the manufacturer. A source of error for microphones is that their frequency response isn't entirely flat, i.e. they pick up some frequencies stronger than others. The microphones, or the structure that they are mounted to, could also resonate at a certain frequency, which will alter the frequency responses. Furthermore, some noise could be induced when the signal is transported back into the PC, and the finite numerical precision will of course always be present when doing numerical calculations.

It is also very important that the gain of the recording microphones is set such that as much as possible of the dynamic range of the A/D converter is used, i.e. that the loudest sound recorded is close to the maximal output

of the A/D converter. This will assure that the available number of bits in the A/D converter is used optimally, providing better resolution in the output signal.

2.4 The Parabolic Reflector

It was decided to use parabolic reflectors to change the directional properties of the microphones for these experiments. Another possible solution would have been to use an interference tube, like in a shotgun microphone. However, using a reflective structure offers more flexibility in the design, and the interference tube has some properties that are undesirable for these experiments. An interference tube can suppress sounds coming from its sides, but is likely to be sensitive to sound coming from behind [23] [24]. A typical interference tube is between 20 and 30 centimetres long, making it too big to be fitted in any portable communication device. Compared the interference tube, the parabolic reflector amplifies sound coming from a certain direction rather than suppressing sounds coming from other directions like the shotgun microphone.

A parabolic reflector was chosen since it focuses the incoming waves onto a single point, thus providing the biggest possible amplification [25]. Furthermore the parabolic reflector delays all waves reflected to the focal point equally [26]. This will produce at least one, but probably several distinct zeros in the frequency response for the microphone. The fact that the delay always is equal for sounds coming directly from the front of the reflector might be used to determine the direction of an incoming sound. If the reflected wave has this specific delay, it is likely to have originated from a source in front of the reflector.

A parabolic reflector focuses sound coming in parallel to the horizontal symmetry line of the parabola onto a single point, the focal point. Sound coming in from other angles are reflected away from the focal point. As illustrated in Figure 2.7. Placing a microphone in the focal point of the parabolic reflector will yield the result that sound coming parallel to the horizontal symmetry line is amplified and reflected into the microphone. Sound coming from other directions is either unaffected or reflected away from the microphone.

The dimensions describing the reflector are shown in Figure 2.8. Due to the properties of a parabolic reflector the time difference to the focal point between the direct and reflected wave is constant, the additional distance the wave has to travel is $2L$, where L is the focal length of the parabolic reflector. The focal length L is related to the diameter D and depth d as

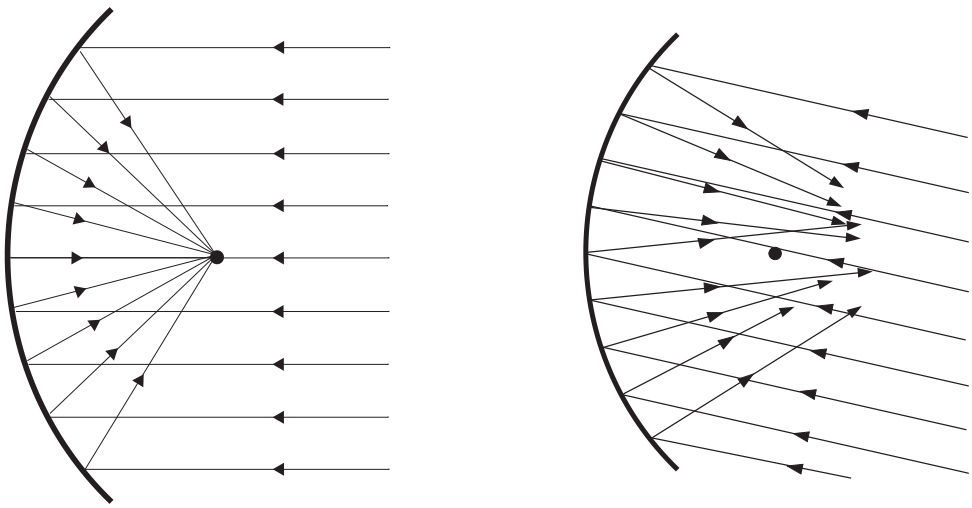

Figure 2.7: An illustration of a parabolic reflector. The parabolic reflector focuses waves coming in parallel to its horizontal symmetry line to a single point. Waves coming in from another direction are not focused.

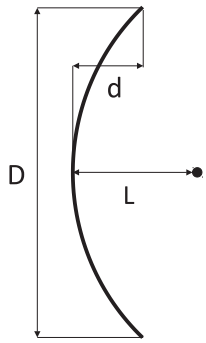

Figure 2.8: The dimensions of a parabolic reflector.

$L = \frac{D^2}{16d}$ [26]. The gain from a parabolic reflector is given by

$$G = \eta(\frac{\pi D}{\lambda})^2 \tag{2.33}$$

where η is the reflector efficiency, determined by the material and construction properties of the reflector and λ is the wavelength of the incoming sound [26].

2.4.1 Sizing

Due to the wave characteristics of sound, a reflecting surface will act as a high pass filter [27]. If the wavelength is larger than the diameter of

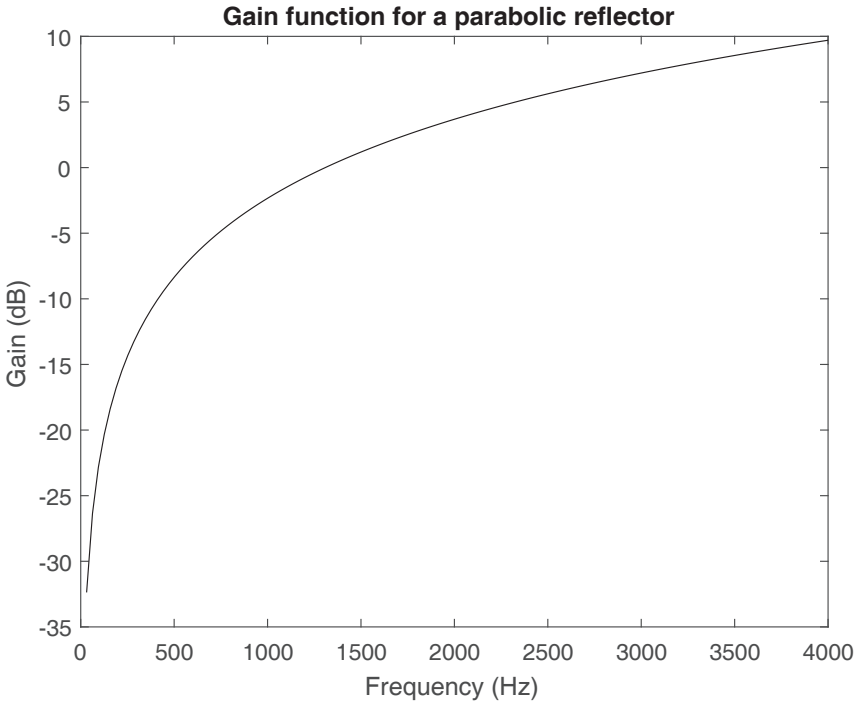

Figure 2.9: The gain function for a parabolic reflector with diameter 10 cm and $\eta = 0.7$.

the reflecting surface the greater part of the sound pressure will diffract around the object rather than being reflected. The proportion between reflected and diffracted sound pressure depends on the structure and material properties of the reflecting object, but if the object is less than 1/3 of the wavelength, there will be virtually only diffraction [28].

For this application we are primarily interested in separating speech from noise. In a few different sources there are some disagreements of where the human voice frequency range should be defined, but it seems to be in the range of 300-5000 Hz [29] [30]. Where the higher frequencies generally correspond to consonant sounds [31].

The idea of this thesis is that a reflecting structure could be built into for example hearing aids or portable communication devices, so it should be made as small as possible, but there will be a lower limit in size where the sound just diffracts around the reflector rather than being reflected. Where this limit is will be investigated, and the result will be used to determine if there is potential for the usage of reflective structures in small devices.

In [32], a rule of thumb for the lowest frequency where a circular reflector

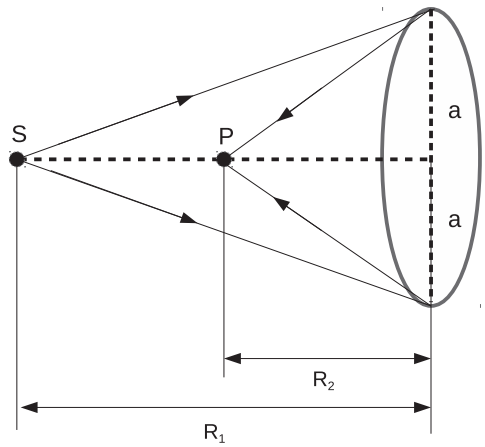

Figure 2.10: Sound reflection on a circular disc.

is effective defined as:

$$f_{min} = \frac{c\tilde{R}}{4a^2} \approx 85\frac{\tilde{R}}{a^2}\ Hz \tag{2.34}$$

where a is the radius of the circular reflector and c is the speed of sound. The harmonic mean distance $\tilde{R}$ is

$$\tilde{R} = 2(\frac{1}{R_1} + \frac{1}{R_2})^{-1} \tag{2.35}$$

where R_1 is the distance between a point source and the center of the disc and R_2 is the distance between the observation point and the center of the disc, as shown in Figure 2.10 . The effective frequency is defined in [32] as the frequency where the reflected sound pressure is half of the maximum possible reflected sound pressure.

Frequencies below the minimum efficient frequency will diffract around the reflector rather than being reflected back. According to Figure 2.11, this means that to effectively reflect the lowest frequency in the voice band, 300 Hz, the reflector would have to be almost 50 cm in diameter. This is too big to be fitted in any portable communication device. A reasonable biggest reflector diameter to test is 10 cm, considering both the portability aspects, and the limitations of the 3D-printer that will be used to manufacture the reflector.

The goal of the reflector here is to make one of the channels sufficiently different from the others so that there is additional diversity in the mixing matrix. How much diversity is required to notably improve the source

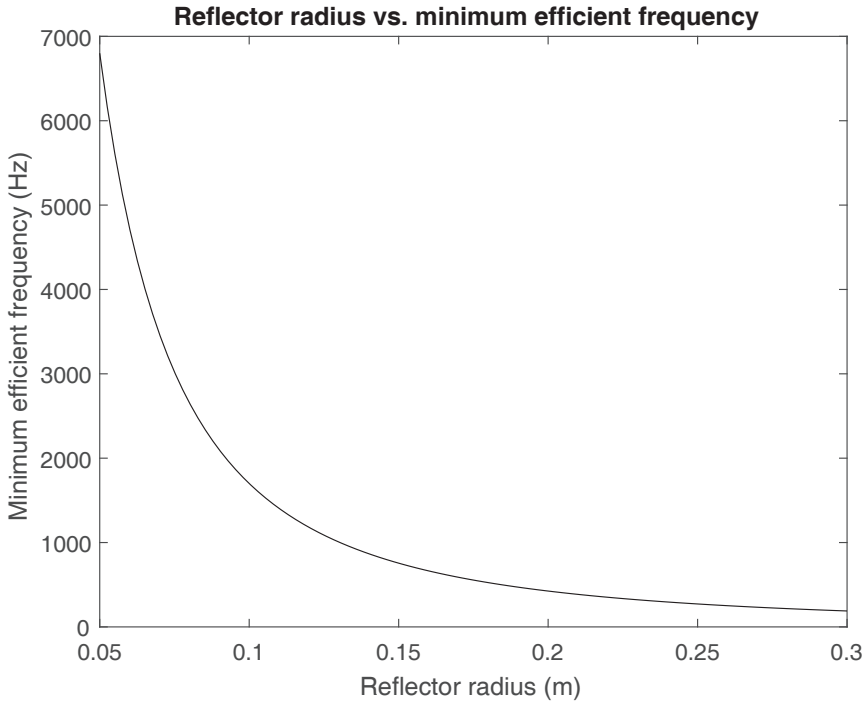

Figure 2.11: Circular reflector radius vs. minimum efficient frequency.

separation, and all the ways in which ways a reflector adds diversity is difficult to simulate, and measurements using differently sized reflectors will have to be made. Starting at a large size, and then trying smaller sizes. It could be that a changed frequency response for higher frequencies is sufficient to improve the source separation.

The parabolic reflector will have a circular hole in the middle. The reason for this is to allow the sound coming in from the front to enter the second microphone unobstructed by the reflector. This hole will cause diffraction of the incoming signal, however, since the second microphone is placed on a horizontal axis passing through the center point of this hole, the diffraction effects experienced by the second microphone are small [25].

2.4.2 Microphone distance

The purpose of using a reflective structure for one of the microphones is to make this channel behave differently for sound coming in parallel to the symmetry line of the reflector. For sound coming from another direction the two microphones should behave as similar as possible, i.e. sound coming in

at an angle that isn't parallel with the reflectors symmetry line should appear similarly regardless of the incoming angle. This means that the second microphone should be placed so that it is as close to the first microphone as possible, without being noticeably affected by the reflector. A small phase difference between the signals should result in the signals received by the two microphones being fairly similar, but the interesting signal will be stronger in the microphone which has a reflector. It was decided to place the microphones 5 cm apart for this experiment since the second microphone then had an unobstructed line of sight for almost all directions, which means that it shouldn't be notably affected by the reflector.

2.4.3 The reflectors used in the experiments

Trying to maximise the determinant based on the diameter of the reflector will only yield that bigger is better, since a bigger reflector increases the amplification at microphone 1, which will increase the determinant. The depth of the reflector is the second design parameter of the reflectors. The depth determines the focal length of the reflector. The ratio between the focal length L and the diameter D defined by L/D, should be between 0.5 and 0.6 to make the parabolic reflector less sensitive to geometrical errors during manufacture [26]. However a dish that is deeper has better directional properties than a shallow one [25]. The reflectors used in these experiments were therefore designed to be as deep as possible, considering that it should be possible to place the microphone in the focal point using the microphone holder.

The biggest reflector is 10 cm in diameter and 2.8 cm deep, the second largest is 5 cm in diameter and 1.7 cm deep, and the smallest one is 2.5 cm in diameter and 1.25 cm deep. They have L/D-ratios of 0.28, 0.34 and 0.5 respectively.

2.5 Wiener solution for inverting channels

A method to obtain the optimal inverse FIR-filters in the sense of minimizing the sum of the squared output error is proposed in [15]. Assuming that the mixing filters and the cross correlations of the sources are known, the Wiener solution of the separation filters can be calculated. The separating filters can be found by solving

$$\begin{pmatrix} \mathbf{R}_{\mathbf{x}_1\mathbf{x}_1} & \mathbf{R}_{\mathbf{x}_1\mathbf{x}_2} \\ \mathbf{R}_{\mathbf{x}_2\mathbf{x}_1} & \mathbf{R}_{\mathbf{x}_2\mathbf{x}_2} \end{pmatrix} \begin{pmatrix} \mathbf{w}_{11} & \mathbf{w}_{12} \\ \mathbf{w}_{21} & \mathbf{w}_{22} \end{pmatrix} = \begin{pmatrix} \mathbf{r}_{\mathbf{x}_1\mathbf{s}_1} & \mathbf{r}_{\mathbf{x}_1\mathbf{s}_2} \\ \mathbf{r}_{\mathbf{x}_2\mathbf{s}_1} & \mathbf{r}_{\mathbf{x}_2\mathbf{s}_2} \end{pmatrix} \tag{2.36}$$

where $\mathbf{R}_{\mathbf{x}_i\mathbf{x}_j}$ is the correlation matrix of the mixed signals, which is defined as

$$\begin{gathered}\mathbf{R}_{\mathbf{x}_i\mathbf{x}_j}(kl) = \gamma_{x_ix_j}(k-l) = \\ \sum_{n=-\infty}^{\infty}\sum_{o=-\infty}^{\infty} a_{ij}(n)a_{ij}(o)\gamma_{s_is_j}[(k-l-n+o)] \\ (k,l) \in \{Z^2 : |k-l| < L\},\ (i,j) \in \{1,2\}\end{gathered} \tag{2.37}$$

for entry (k,l) and $\gamma_{x_ix_j}$ denotes the cross correlation function of the zero mean mixed signals. The integer L is the length of the separation filter that is to be calculated.

The FIR separation filters that are to be calculated are arranged as column vectors

$$\mathbf{w}_{ij} = [w_{ij}(0)\ w_{ij}(1)\ ...\ w_{ij}(L-1)]^T \tag{2.38}$$

The correlation vectors are defined as column vectors for entry k as

$$\begin{gathered}\mathbf{r}_{\mathbf{x}_is_i}(k) = \gamma_{x_is_j}(k-D) = \sum_{s=-\infty}^{\infty} a_{ij}(n)\gamma_{s_is_j}(k-n-D) \\ (k = 0,\ 1,\ ...,\ L-1)\end{gathered} \tag{2.39}$$

Then, equation 2.36 can be solved by rewriting it as

$$\mathbf{R}\mathbf{w} = \mathbf{r} \iff (\mathbf{I} \otimes \mathbf{R})\,\mathrm{vec}(\mathbf{w}) = \mathrm{vec}(\mathbf{r}) \tag{2.40}$$

where $\otimes$ is the Kronecker product, $\mathbf{I}$ is the 2-by-2 identity matrix, and $\mathrm{vec}(\cdot)$ is a single column vector of all concatenated column vectors of the argument.

2.6 Perceptual evaluation of speech quality (PESQ)

PESQ is a family of standards regarding a test methodology for the evaluation of speech quality experienced by a user of a telephony system. It is a worldwide applied industry standard for objective speech quality evaluation used by phone manufacturers, network equipment vendors and telecom operators.

PESQ compares an original signal $x(t)$ with a degraded signal $y(t)$ that is the result of $x(t)$ being passed through some sort of filter or communications system. The output of PESQ is a numerical score predicting the perceived quality that would be given to $y(t)$ in a subjective listening test. The range of the PESQ score is between -0.5 and 4.5, although for

most cases the output range will be a listening quality mean objective score (MOS) between 1.0 an 4.5 [33].

In the first step of the PESQ evaluation, the delay between the original input and the degraded signal is determined. A series of delays between the original input and the degraded signal are computed, one for each time interval for which the delay is significantly different from the previous time interval. For each of these intervals a corresponding start and stop point is calculated. The alignment algorithm then tries to determine which is the real delay between the original and degraded signal.

Using the set of delays that are found PESQ compares the original signal with the aligned degraded output. To do this comparison both the original and the degraded signal are transformed into an internal representation that is analogous to the psychophysical representation of audio signals. The internal representation is calculated in several stages including time alignment, level alignment to a calibrated listening level, time-frequency mapping, frequency warping and compressive loudness scaling.

The internal representation is processed to compensate for effects such as local gain variations and linear filtering that, if not too severe, have little effect on the perceived speech quality. More severe effects, or rapid variations are only partially compensated, so that a residual effect remains and contributes to the overall perceptual disturbance. Many of the steps in PESQ are algorithmically complex, and not easily described by mathematical formulae. Figures 2.12-2.15 give an overview of the algorithm by block-diagrams, and a high-level description for each block is given in [33].

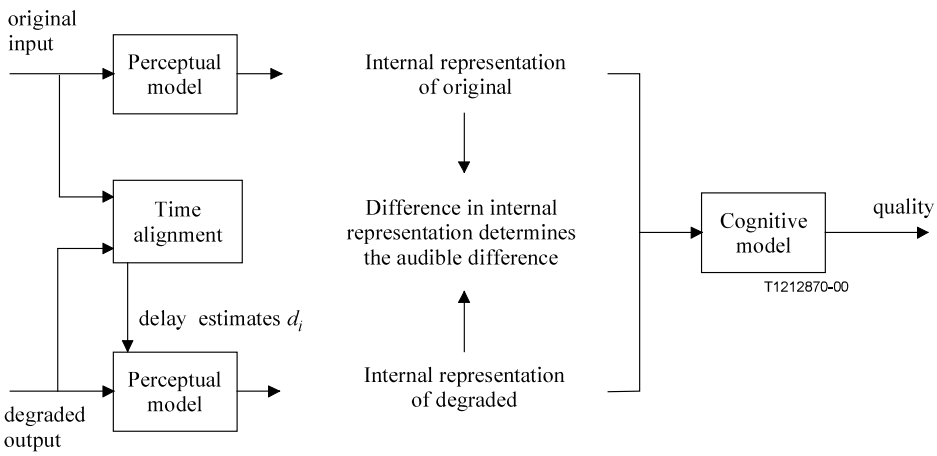

Figure 2.12: An overview of the PESQ evaluation [33].

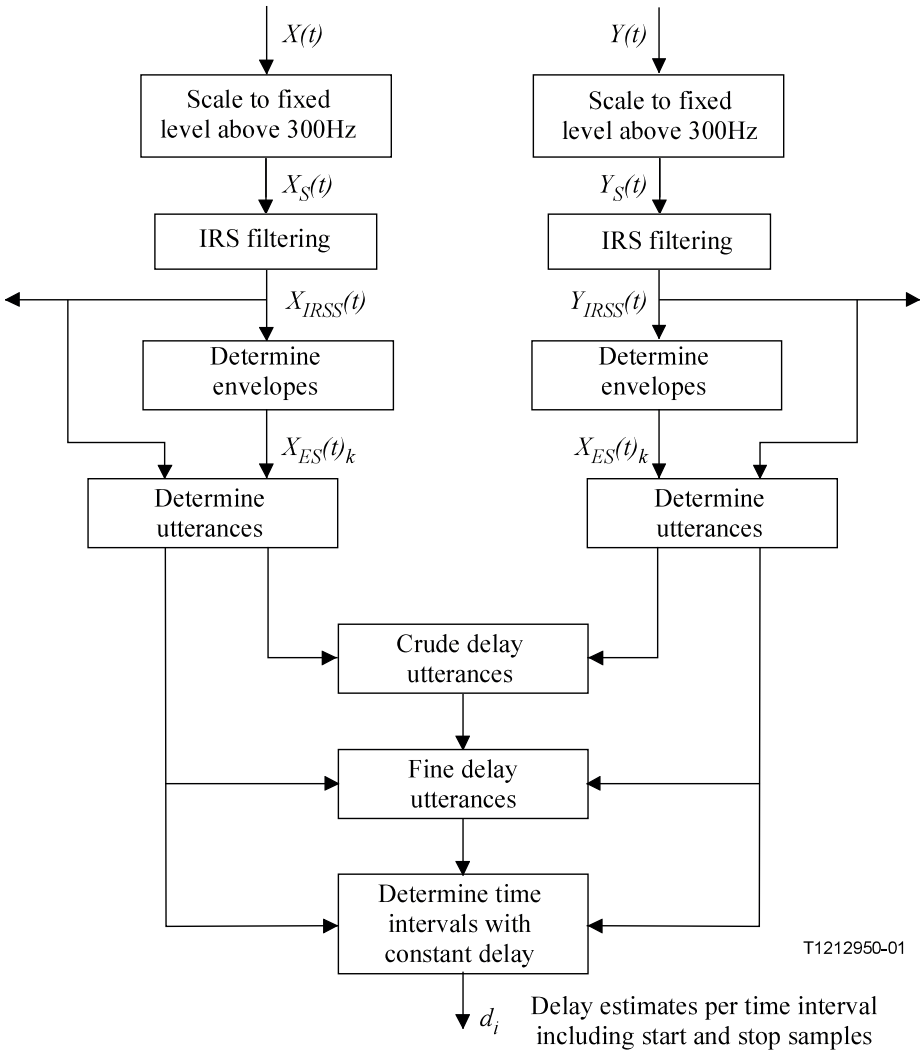

Figure 2.13: The alignment routine used in PESQ to determine the delay per time interval, d_i [33].

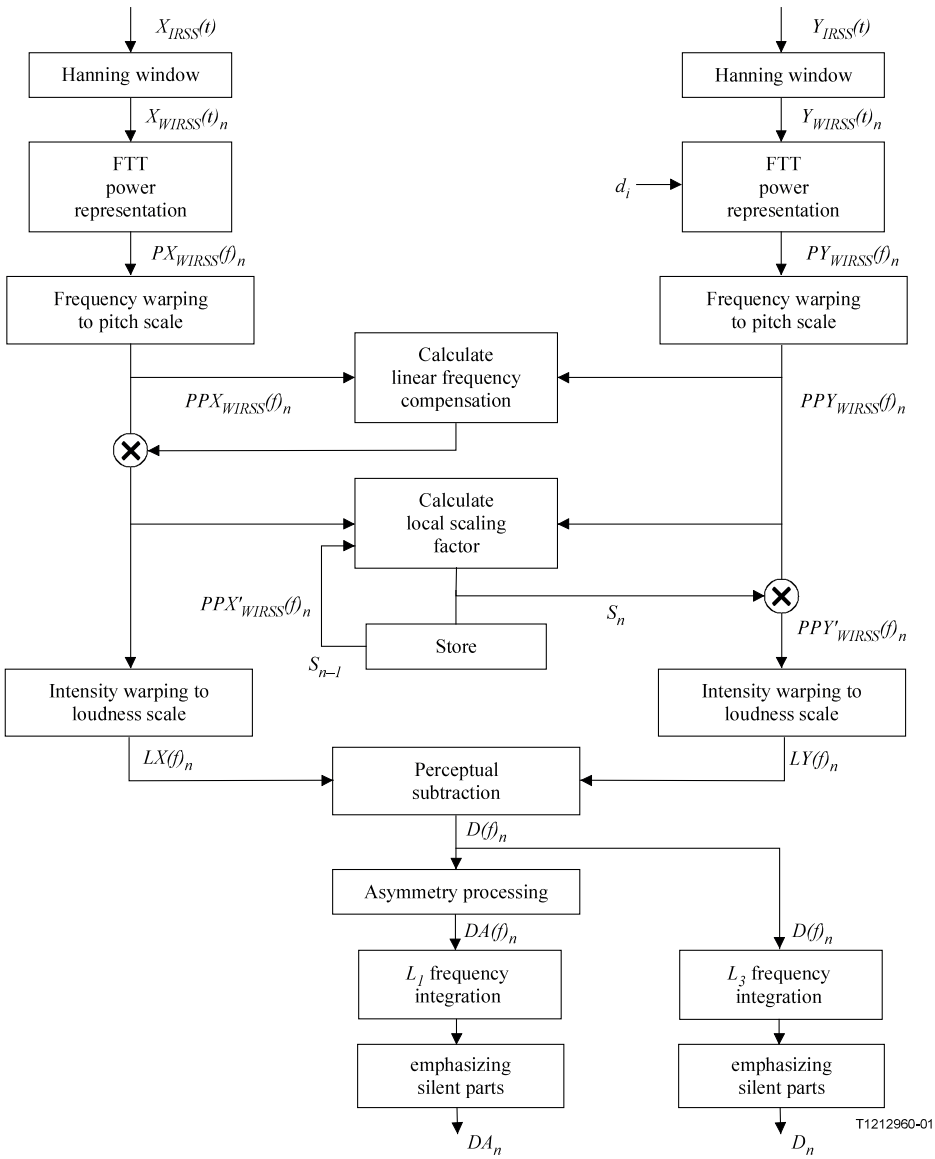

Figure 2.14: Overview of the perceptual model. The distortions per frame D_n and DA_n have to be aggregated over time (index$_n$) to obtain the final disturbances, see Figure 2.15 where the realignment of the degraded signal is given [33].

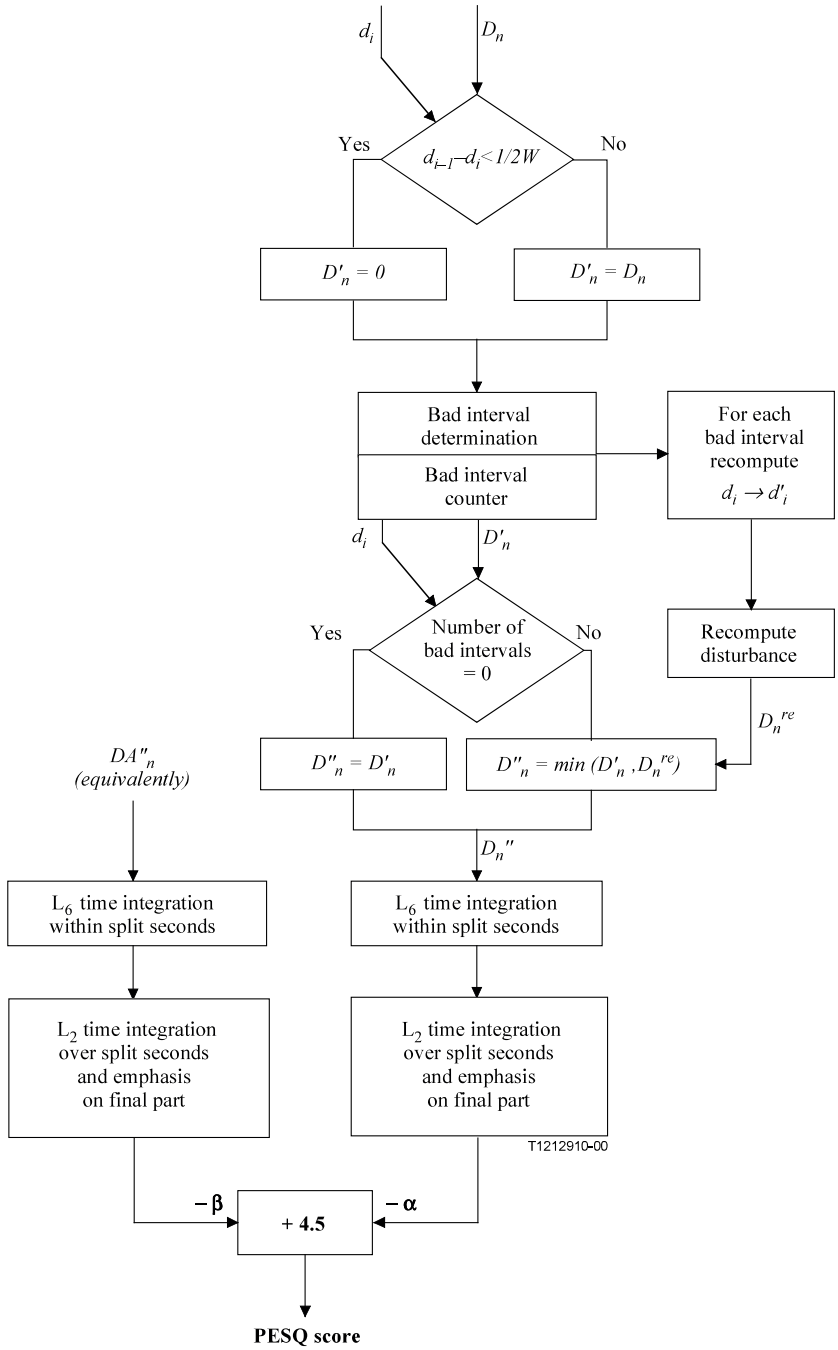

Figure 2.15: After realignment of the bad intervals, the distortions per frame D''_n and DA''_n are integrated over time and mapped to the PESQ score. W is the FTT window length in samples [33].

2.6.1 Signal-to-noise ratio

The signal-to-noise ratio (SNR) is a measure of the relative power between two signals, usually a signal one is interested in and a noise signal. It is defined as

$$SNR = \frac{P_{signal}}{P_{noise}} \tag{2.41}$$

where P is the average power. The ratio is usually expressed in dB as

$$SNR_{dB} = 10 \log_{10} \left(\frac{P_{signal}}{P_{noise}} \right). \tag{2.42}$$

Chapter 3
Simulations

3.1 Simulations of frequency responses and mixing matrix determinant for the parabolic reflector

The frequency response functions for the simulations were calculated using a ray-tracing model to determine the delays caused by reflections in the parabolic reflector, and the optimal gain for a parabolic reflector, equation 2.33. The microphones are represented by single points in the ray-tracing model. The first microphone is placed in the focal point of the reflector and the second one is placed 5 cm away from the first one.

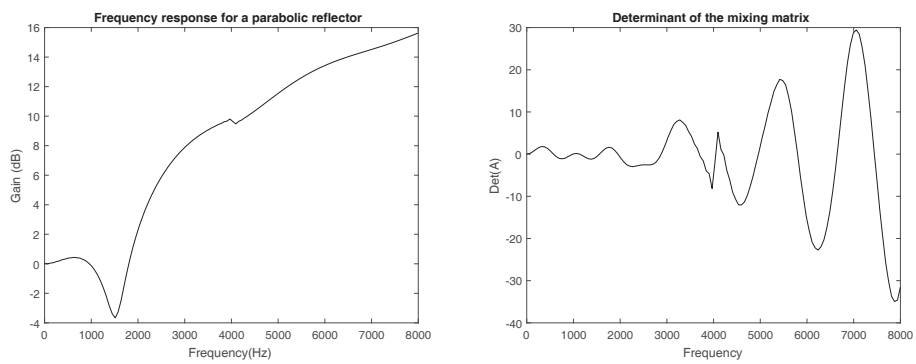

Figure 3.1: The frequency response and mixing matrix determinant for the array with a parabolic reflector with a diameter of 10 cm.

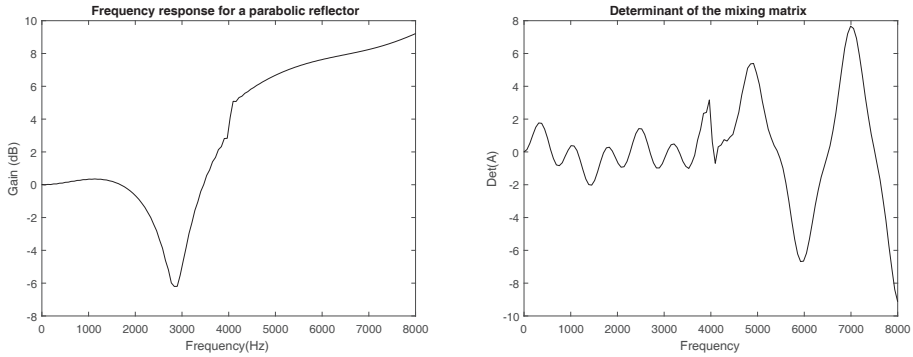

Figure 3.2: The frequency response and mixing matrix determinant for the array with a parabolic reflector with a diameter of 5 cm.

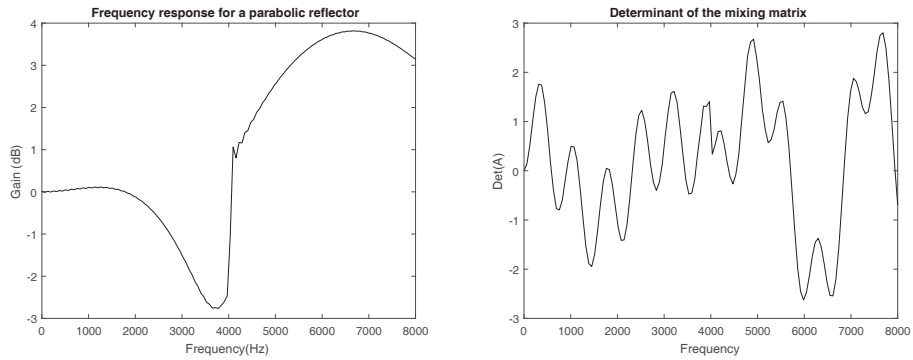

Figure 3.3: The frequency response and mixing matrix determinant for the array with a parabolic reflector with a diameter of 2.5 cm.

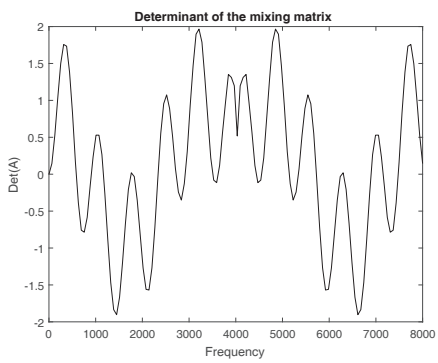

Figure 3.4: The mixing matrix determinant for the array without a reflector.

Chapter 4
Measurements

The parabolic reflector, the rail and microphone holders were designed in SolidWorks and then printed using a EOS Formiga P110 Selective Laser Sintering System. The parts are made using a nylon powder that is melted using a laser. The resolution of this printer is 0.1 mm. The printed rail, reflector and microphone holders can be seen in Figure 4.1.

4.1 Equipment

Two AKG C417 lavalier microphones were used in the array. They are pre-polarized condenser microphones. According to the manufacturer these microphones are omnidirectional and have a flat frequency response up to about 5000 Hz [34].

The Brüel & Kjaer Head and Torso Simulator (HATS) 4128C will be used as one of the sound sources for the measurements, and it will be used to make measurements on the EarIn-headphones when they are mounted in the ears. The HATS mouth is meant to emulate a human mouth and it is mounted inside a cavity in the HATS head [35]. The cavity will cause filtering of the output signal. To compensate for the filtering in the HATS, a reference microphone is placed in a holder in front of the mouth of the HATS. The signal picked up by this microphone is then used as the input signal when doing the H_1 estimation of the channels. An Brüel & Kjaer 4938 microphone connected to a Brüel & Kjaer 2670 pre-amplifier is used as the reference microphone.

A Norsonic 270H connected to a Norsonic 280 Power Amplifier was used as the second sound source. The frequency response of this system can be seen in Figure 4.3. The frequency response plot provided by the manufacturer doesn't show any frequency content above 5 kHz, and it is reasonable to assume that the speaker doesn't play back much at higher frequencies due to the speaker elements being designed like mid-range speakers. Therefore a sampling rate of 16 kHz is sufficient for the measurements.

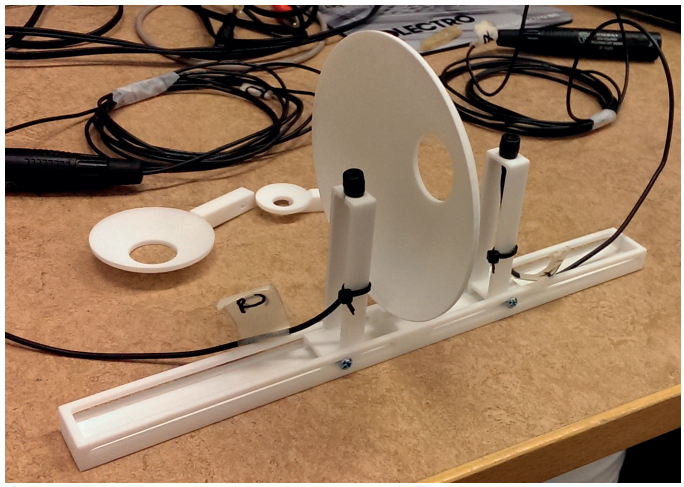

Figure 4.1: The 3D-printed structure.

Microphones are connected to a RME OctaMic XTC microphone preamplifier, which is connected to an RME ADI-8 QS AD/DA-converter, which in turn is connected to a PC running an audio recording software. The microphones will be placed in the 3D-printed rail structure, where the two microphones can be positioned along an axis, and an interchangeable reflective structure can be placed between the microphones.

The frequency response of one of the AKG C417-microphones was measured while it was placed in the array. The gain of the AKG microphone was set to 44 dB and the gain of the reference HATS-microphone was set to 29 dB.

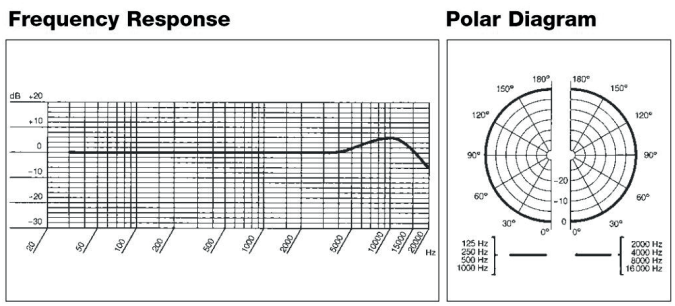

Figure 4.2: The frequency response and the polar pattern of the AKG C417 [34].

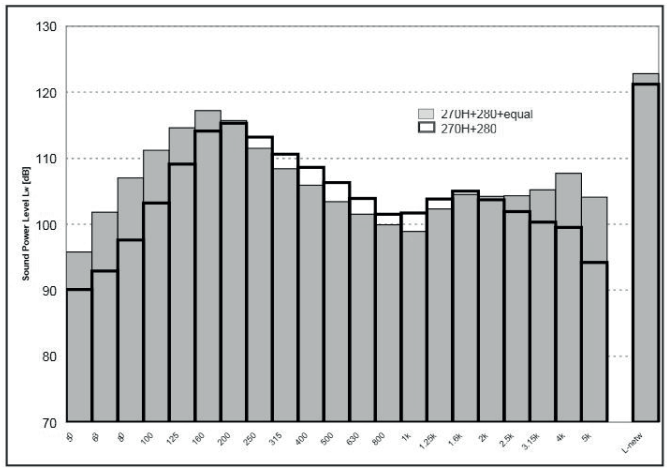

Figure 4.3: The frequency response of the Norsonic 270H Dodecahedron Loudspeaker connected to the Norsonic 280 Power Amplifier [36].

4.2 Polar patterns of the parabolic reflectors

The polar patterns of the reflectors were measured by manually aiming the reflector using a graded paper. The set-up can be seen in Figure 4.5. The measurements were made for 0°, 15° and 30° and then in 30° increments up to 180°. The reference microphone at the mouth of the HATS and one AKG C417 was used, the gain levels were set to 27 dB for the reference microphone and 39 dB for the AKG C417. The coherence function was continuously studied during all of the channel estimations to assure that the measurement noise at the input was low and the spectral resolution was set sufficiently high.

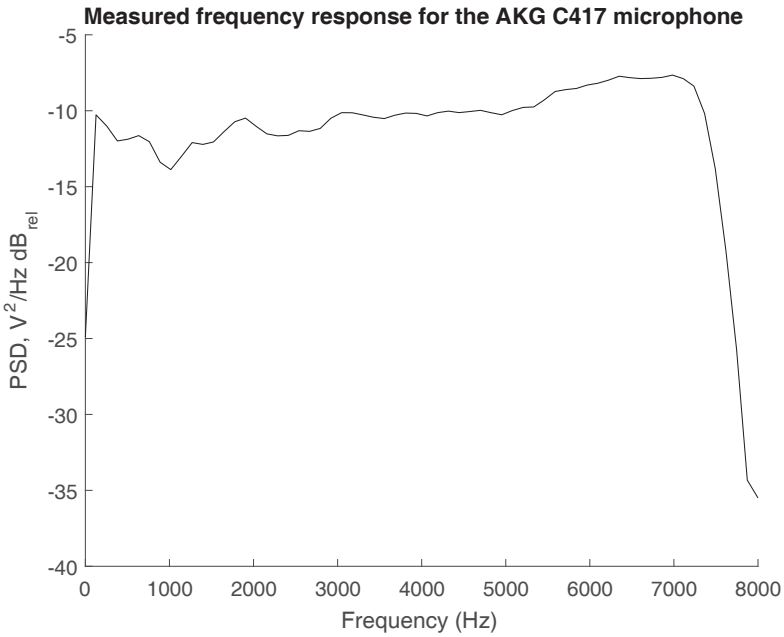

Figure 4.4: The measured frequency response for the AKG C417 microphone which was used in the array.

Figure 4.5: The set-up to measure the polar pattern of the parabolic reflectors.

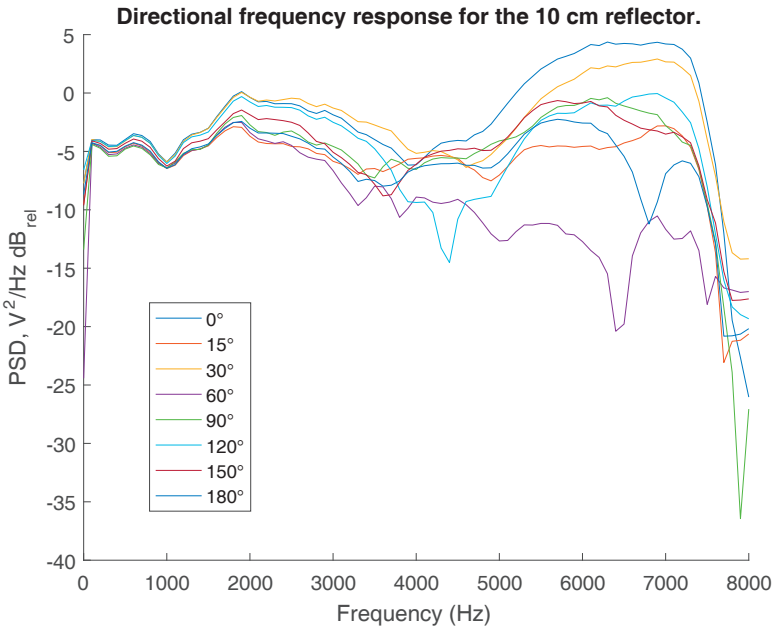

Figure 4.6: The directional frequency response of the 10 cm reflector. The values have been scaled to obtain dB-levels around zero.

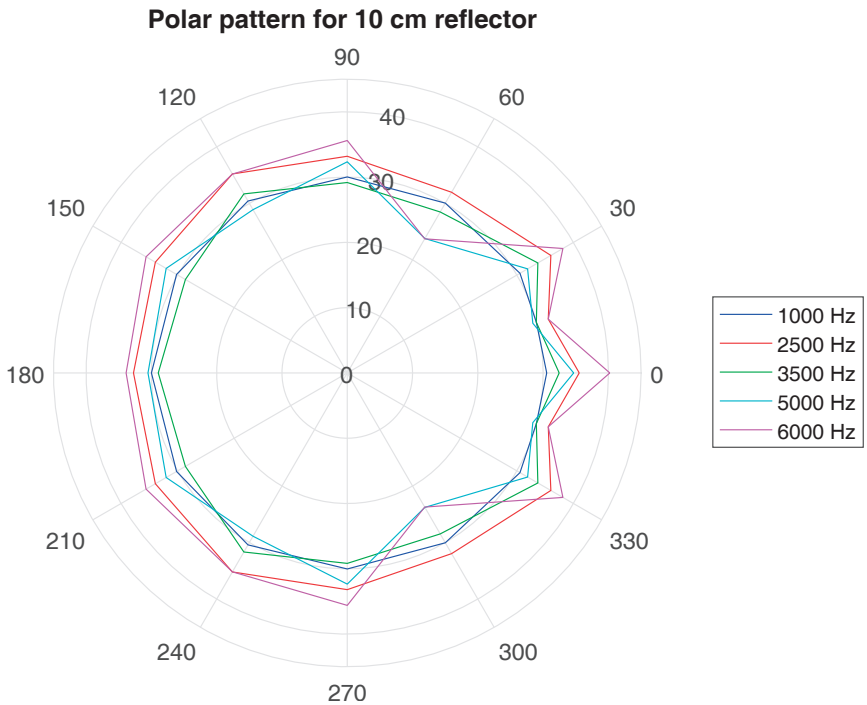

Figure 4.7: The measured polar pattern for the 10 cm reflector.

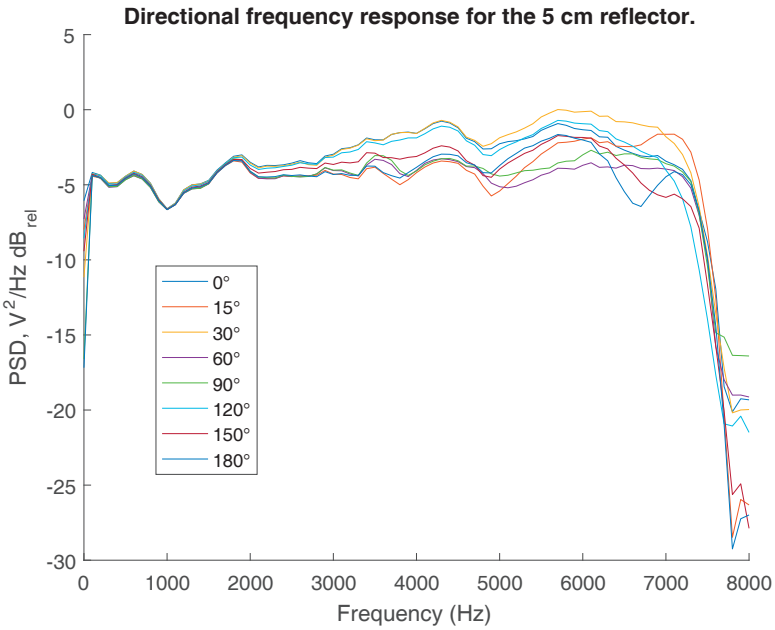

Figure 4.8: The directional frequency response of the 5 cm reflector. The values have been scaled to obtain dB-levels around zero.

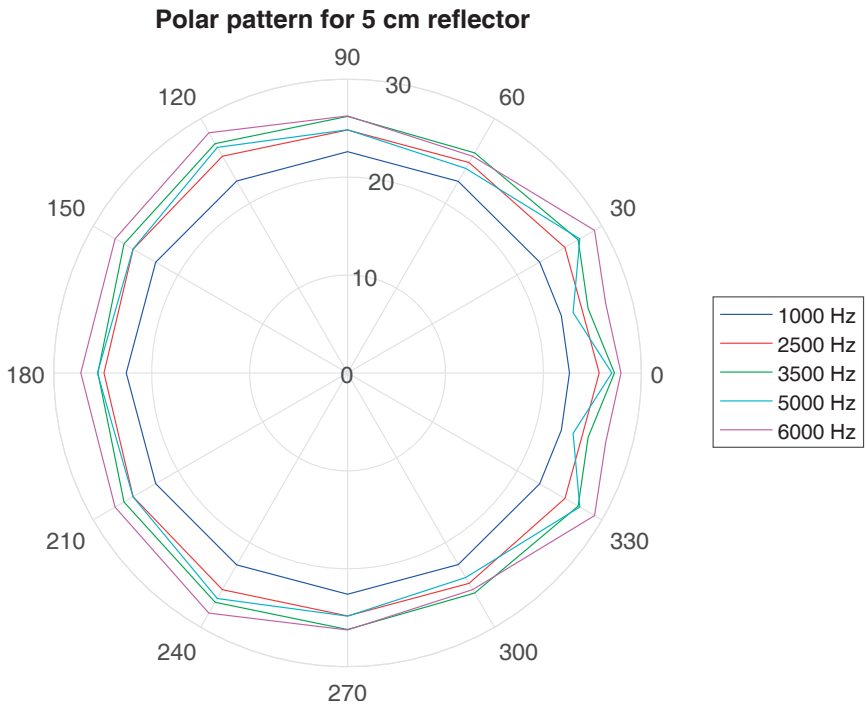

Figure 4.9: The measured polar pattern for the 5 cm reflector.

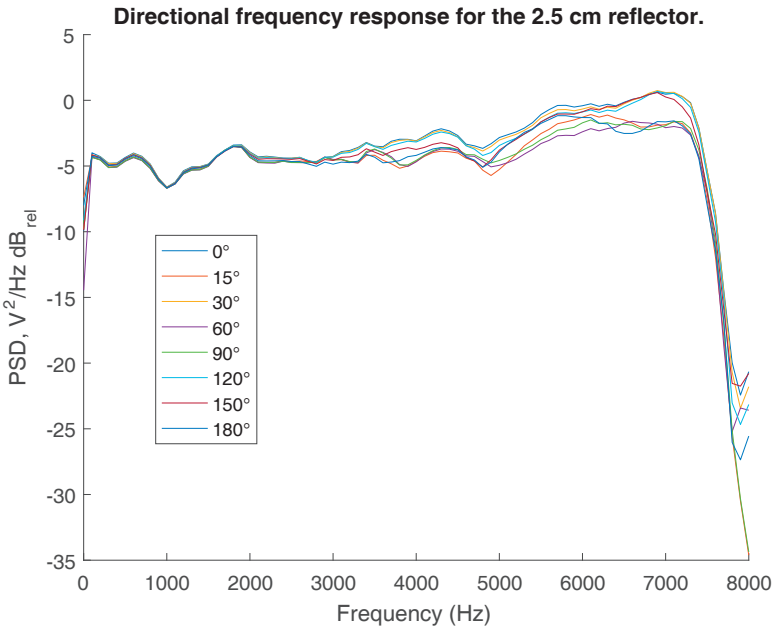

Figure 4.10: The directional frequency response of the 2.5 cm reflector. The values have been scaled to obtain dB-levels around zero.

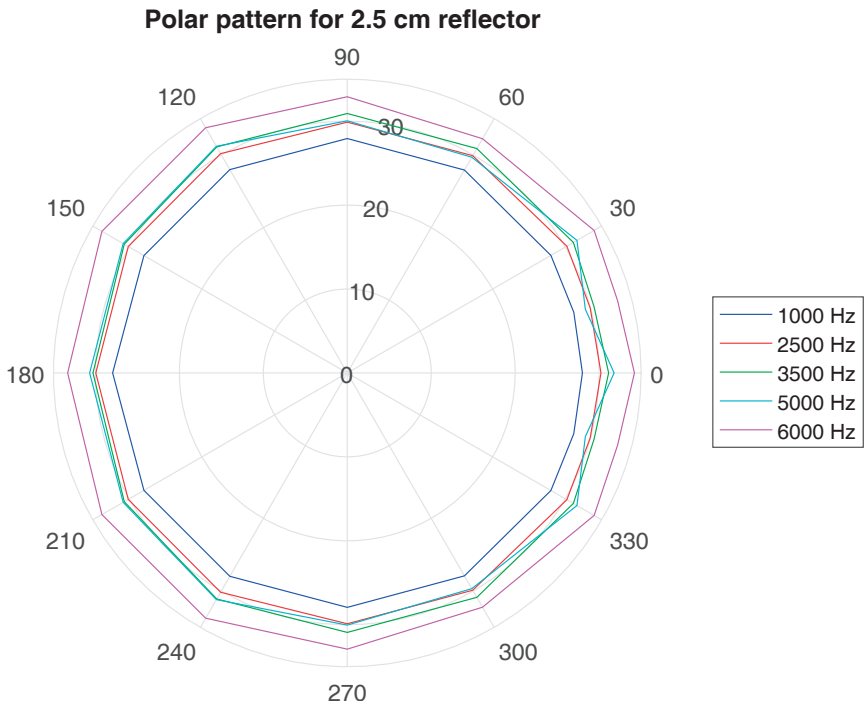

Figure 4.11: The measured polar pattern for the 2.5 cm reflector.

4.3 Frequency responses and mixing matrix determinants

These measurements were made using two sources, the HATS mouth speaker and the Norsonic 270H. Both microphones in the array were used. The sources played back ten seconds of flat-spectrum uncorrelated noise at a sample rate of 16 kHz.

The distance between the HATS mouth speaker and the first microphone was about 1 meter, and the distance between the second microphone and the Norsonic 270H was also about 1 meter. The microphones were placed 5 cm apart, and they were adjusted so that the first microphone was in the focal point the reflector that was used. The gain levels were set to 27 dB for the mouth reference microphone and 39 dB for the two AKG 470C microphones.

4.3.1 Noise behind the array

First, the Norsonic 270H was placed opposite to the HATS, as can be seen in Figure 4.12.

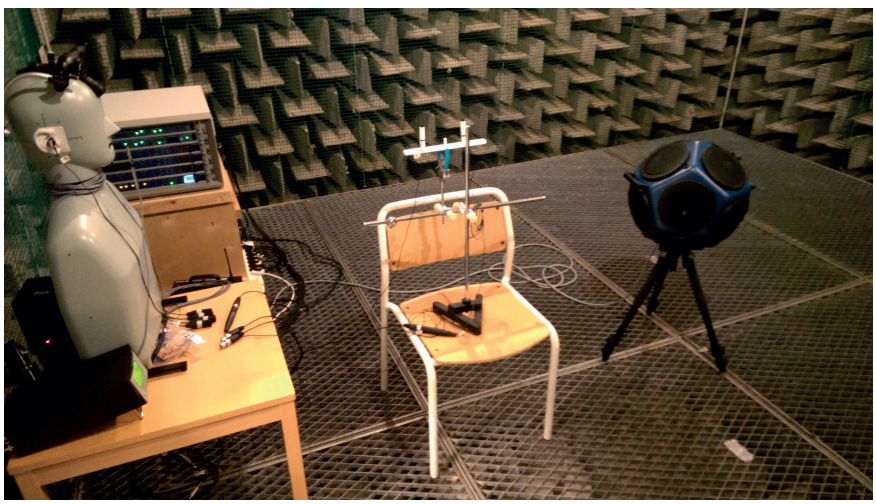

Figure 4.12: The first measurement set-up.

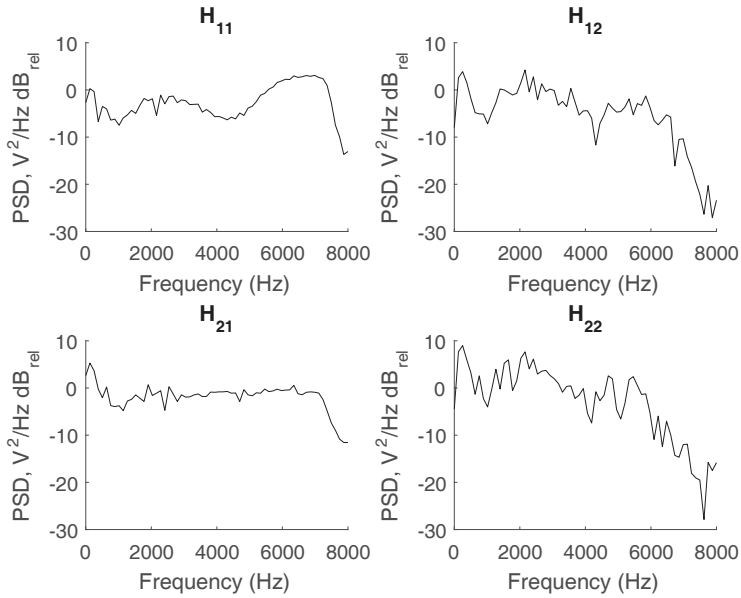

Figure 4.13: The frequency response for the array with a parabolic reflector with a diameter of 10 cm. The noise source is placed behind the array.

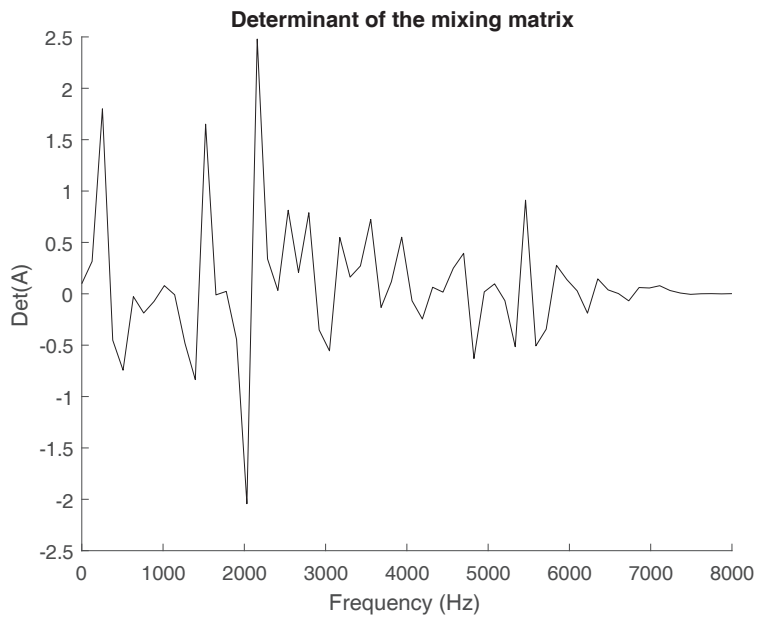

Figure 4.14: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 10 cm. The noise source is placed behind the array.

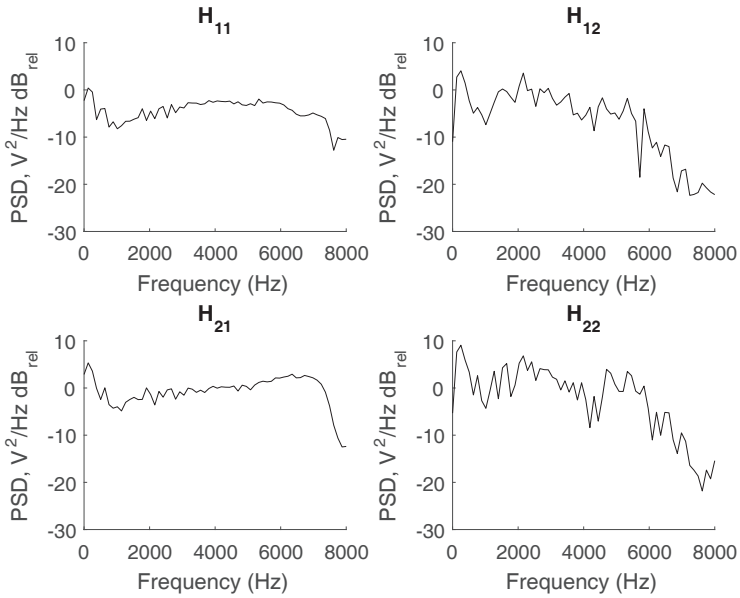

Figure 4.15: The frequency response for the array with a parabolic reflector with a diameter of 5 cm. The noise source is placed behind the array.

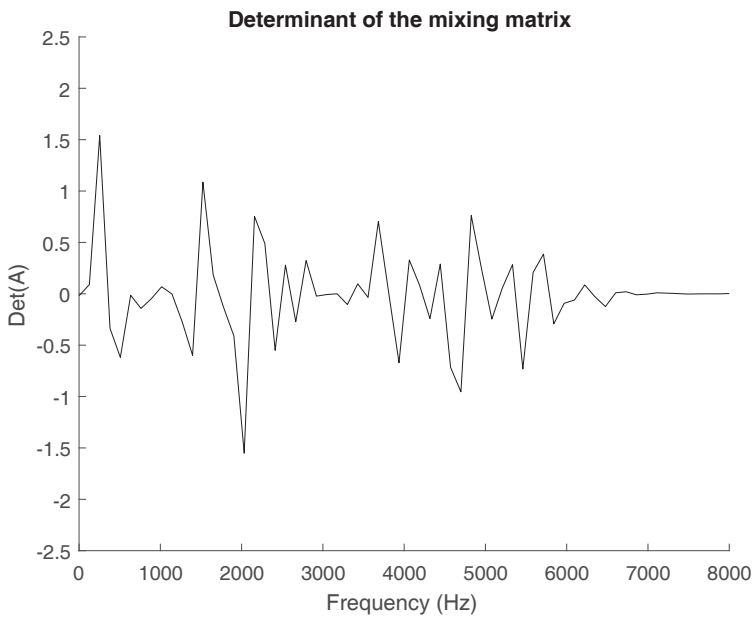

Figure 4.16: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 5 cm. The noise source is placed behind the array.

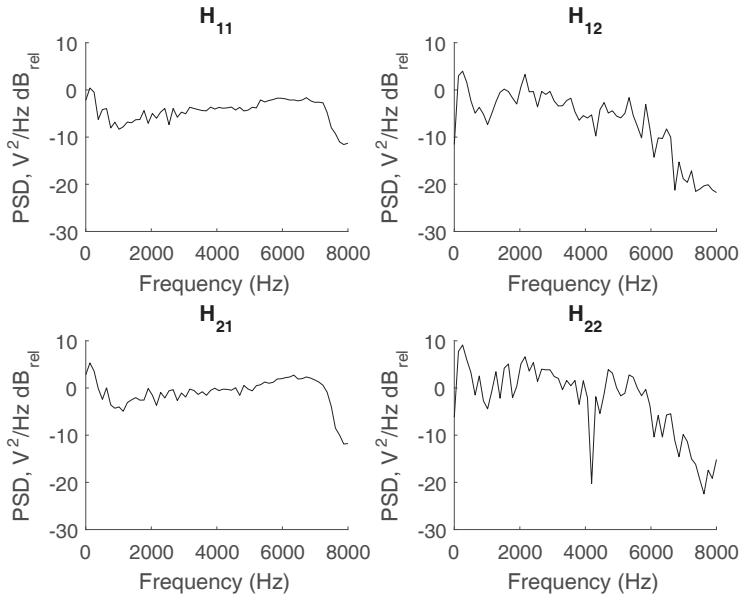

Figure 4.17: The frequency response for the array with a parabolic reflector with a diameter of 2.5 cm. The noise source is placed behind the array.

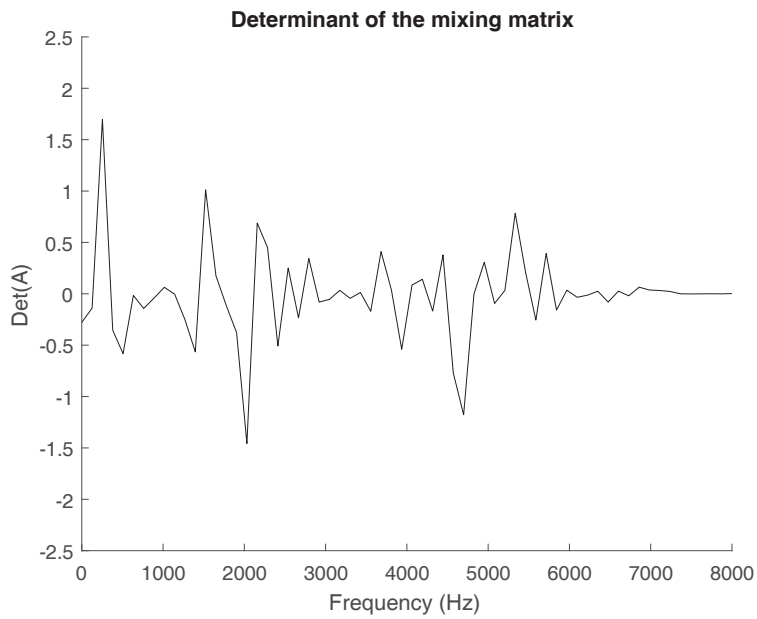

Figure 4.18: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 2.5 cm. The noise source is placed behind the array.

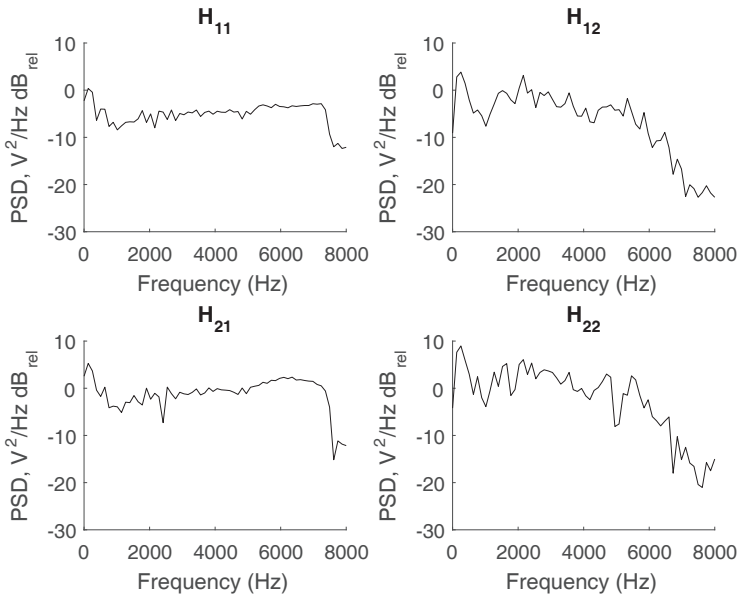

Figure 4.19: The frequency response for the array without a reflector. The noise source is placed behind the array.

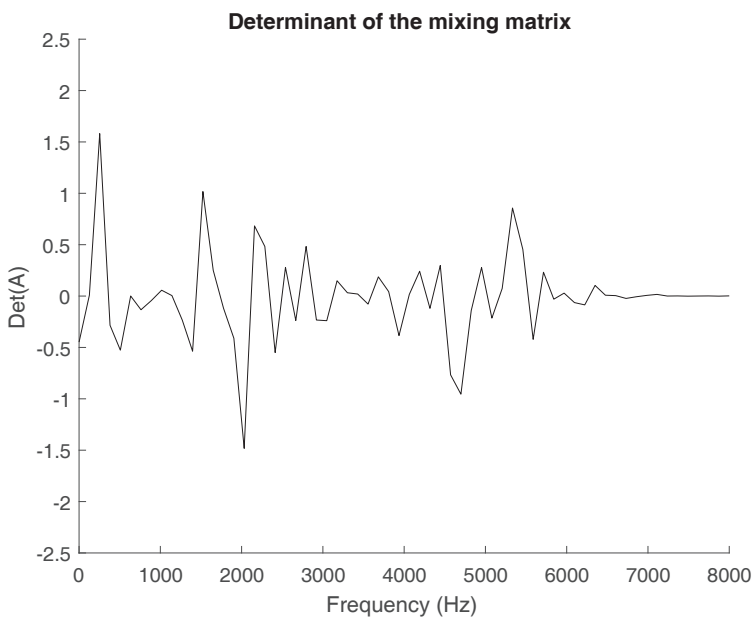

Figure 4.20: The mixing matrix determinant for the array without a reflector. The noise source is placed behind the array.

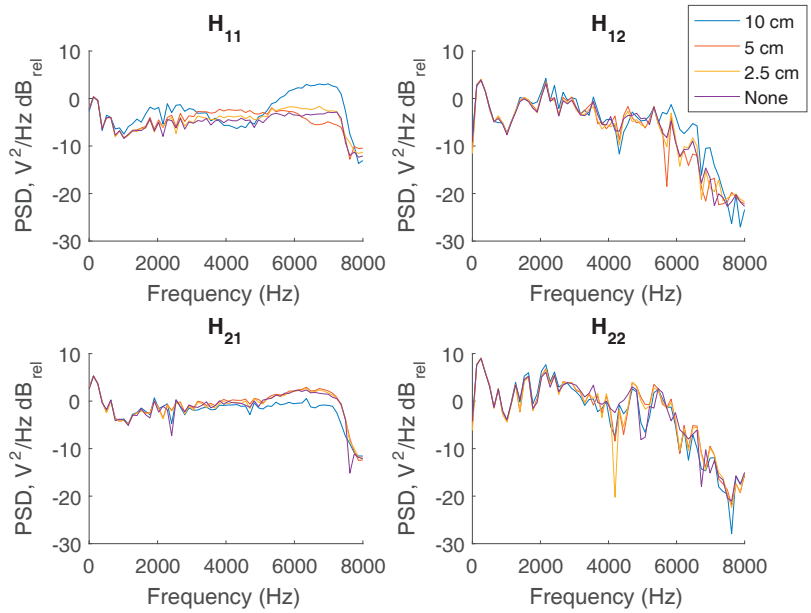

Figure 4.21: The frequency responses for two sources and two microphones when the noise source is placed behind the array.

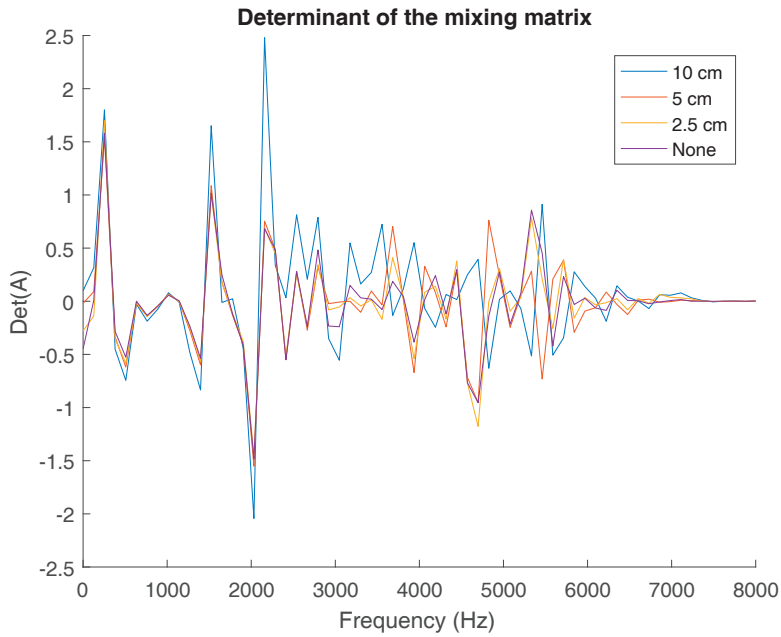

Figure 4.22: The determinants in linear scale of the mixing matrices when the noise source is placed behind the array.

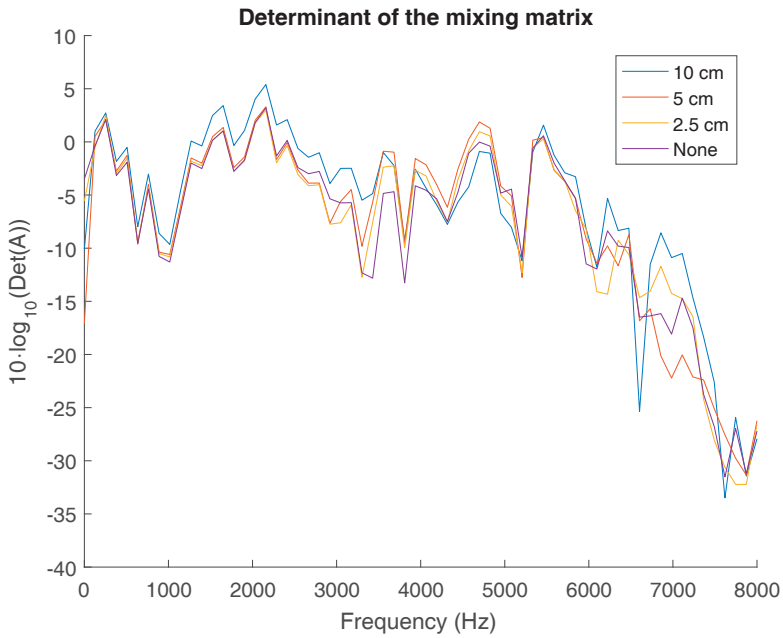

Figure 4.23: The determinants of the mixing matrices in logarithmic scale when the noise source is placed behind the array.

4.3.2 Noise 90° to the right

For the next set of measurements, the Norsonic 270H was placed 90° to the right of the array, as can be seen in Figure 4.24.

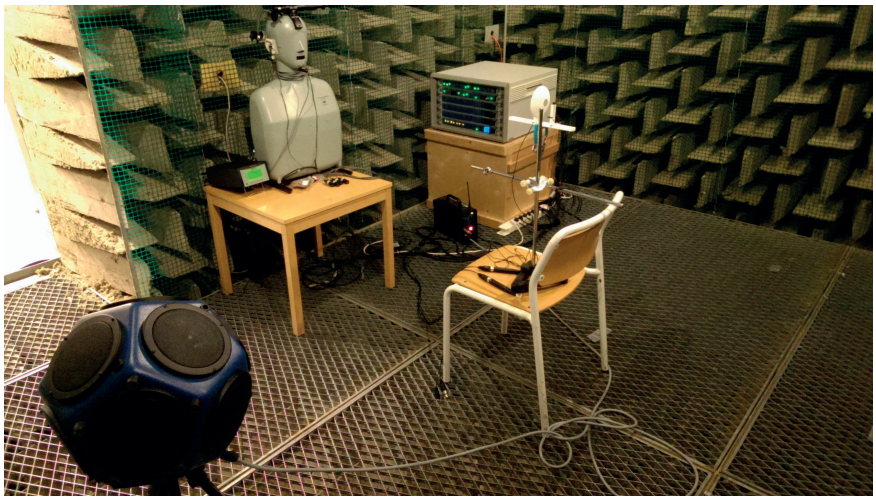

Figure 4.24: The set-up with two sources used to estimate the mixing matrices. The Norsonic 270H was placed 90° to the right of the array.

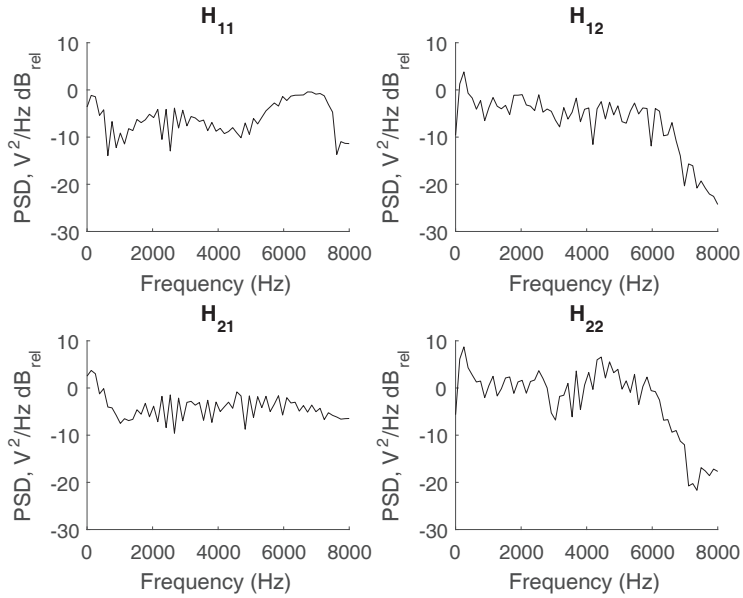

Figure 4.25: The frequency response for the array with a parabolic reflector with a diameter of 10 cm. The noise source is placed 90° to the right of the array.

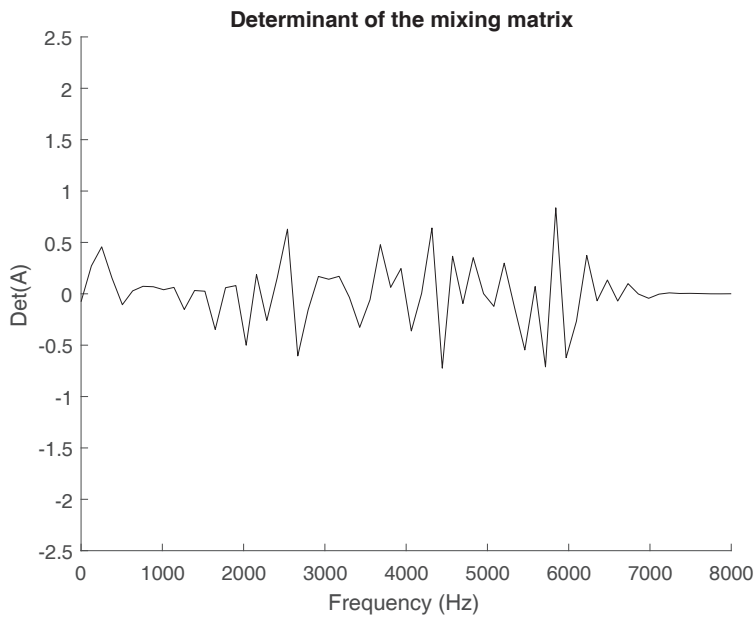

Figure 4.26: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 10 cm. The noise source is placed 90° to the right of the array.

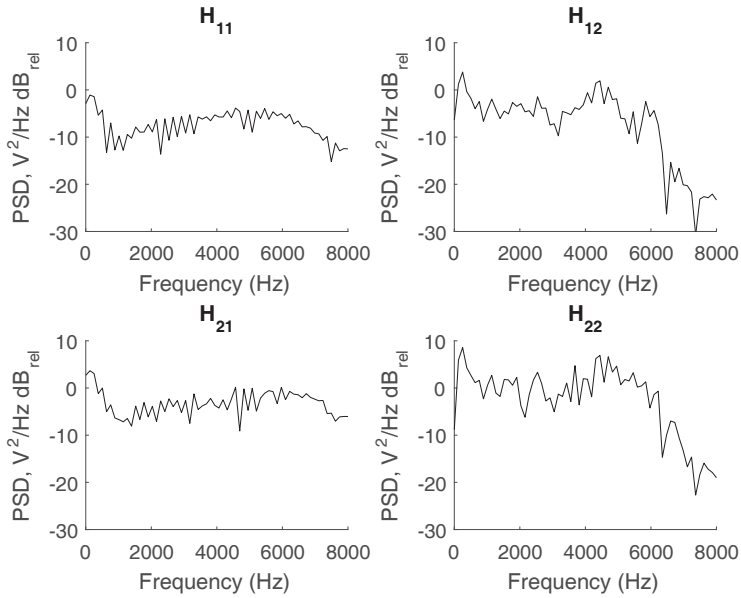

Figure 4.27: The frequency response for the array with a parabolic reflector with a diameter of 5 cm. The noise source is placed 90° to the right of the array.

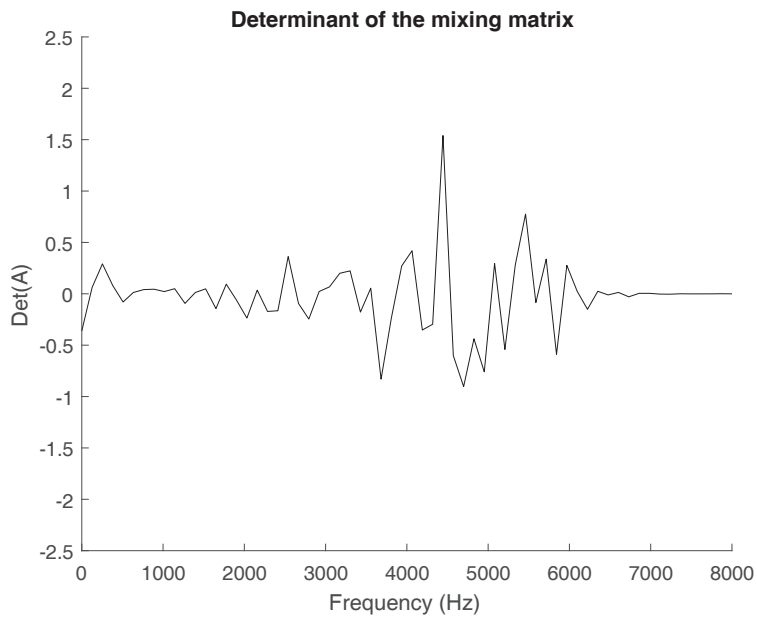

Figure 4.28: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 5 cm. The noise source is placed 90° to the right of the array.

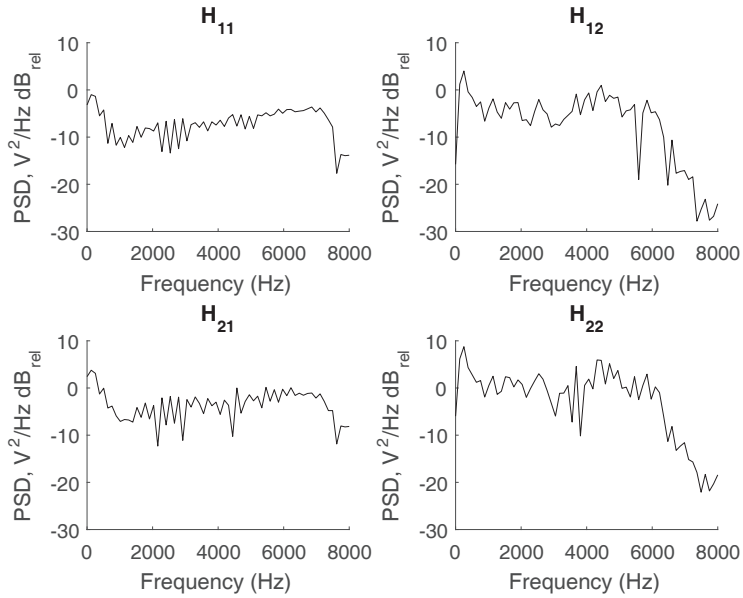

Figure 4.29: The frequency response for the array with a parabolic reflector with a diameter of 2.5 cm. The noise source is placed 90° to the right of the array.

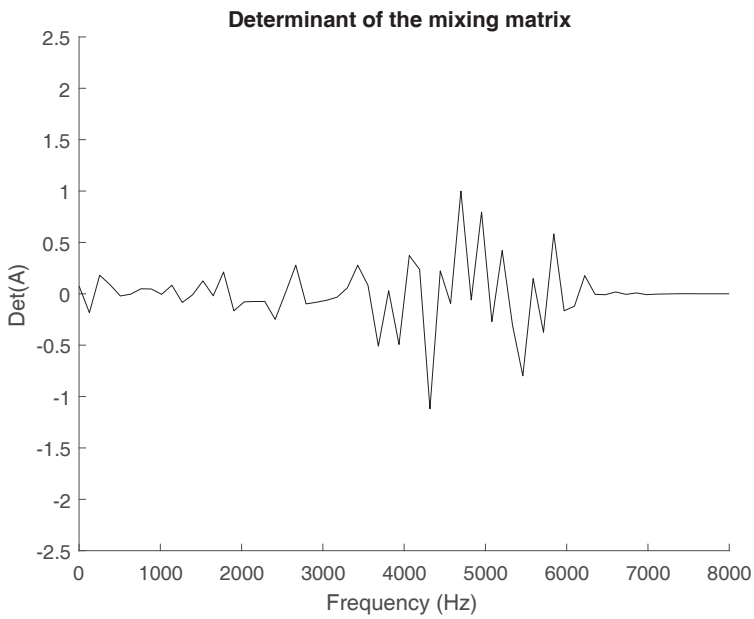

Figure 4.30: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 2.5 cm. The noise source is placed 90° to the right of the array.

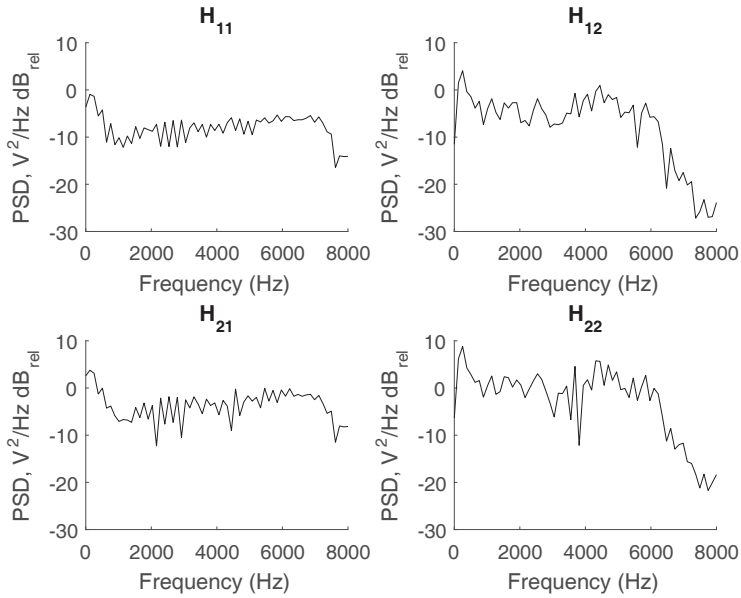

Figure 4.31: The frequency response for the array without a reflector. The noise source is placed 90° to the right of the array.

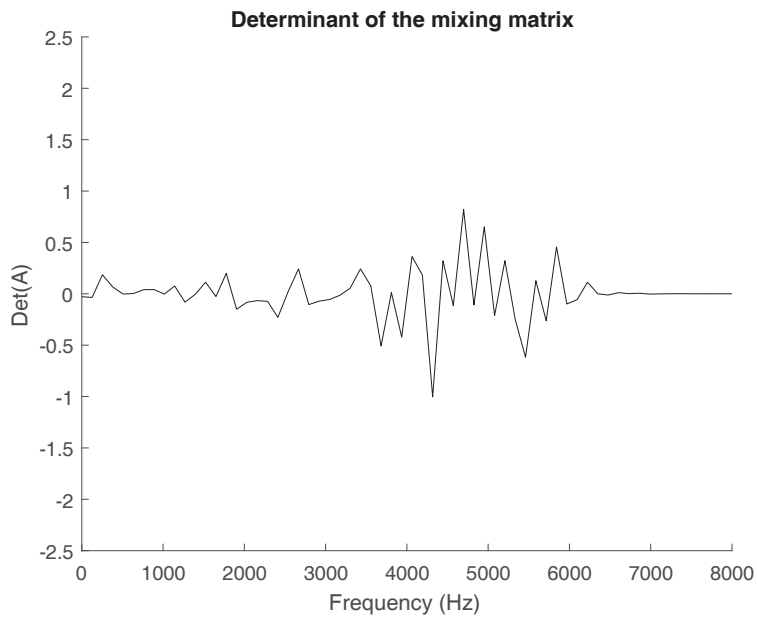

Figure 4.32: The mixing matrix determinant for the array without a reflector. The noise source is placed 90° to the right of the array.

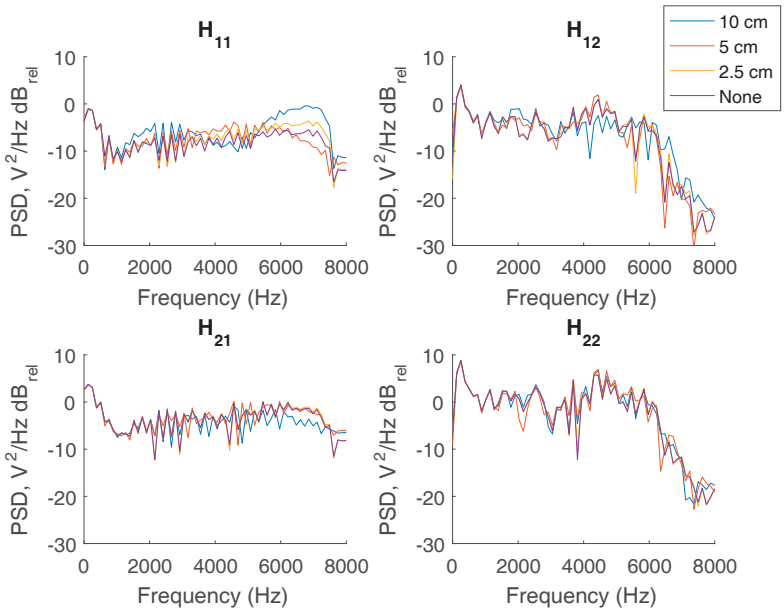

Figure 4.33: The frequency responses for two sources and two microphones with the noise source placed 90° to the right of the array.

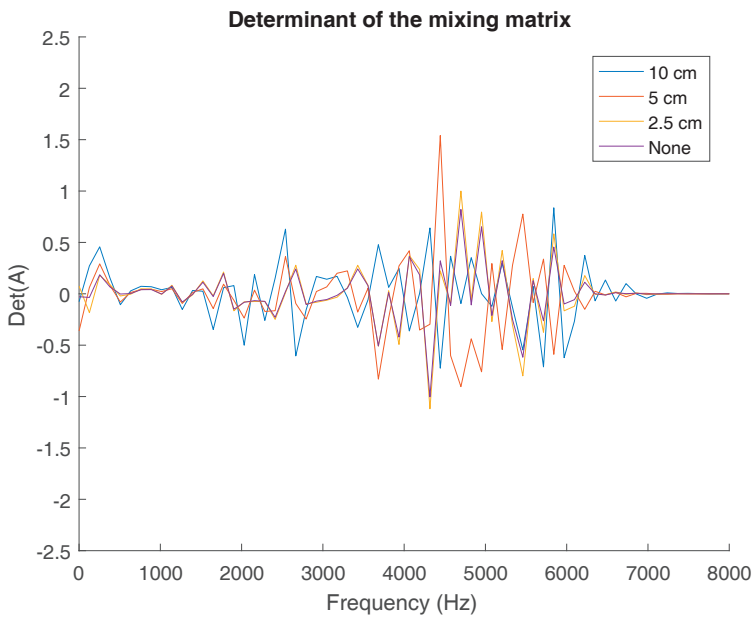

Figure 4.34: The determinants of the mixing matrices in linear scale when the noise source is placed 90° to the right of the array.

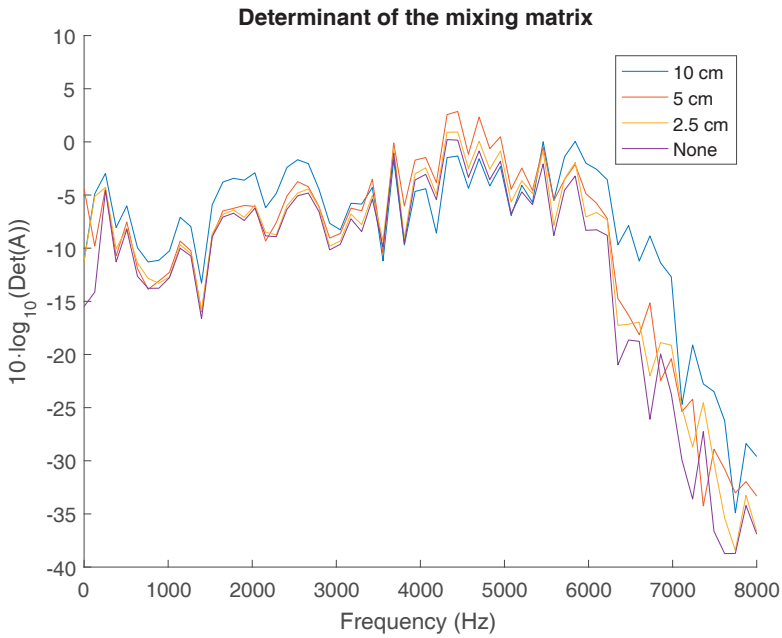

Figure 4.35: The determinants of the mixing matrices in logarithmic scale when the noise source is placed 90° to the right of the array.

4.3.3 Noise in front of the array

Lastly, the Norsonic 270H was placed directly to the right of the HATS, as can be seen in Figure 4.36.

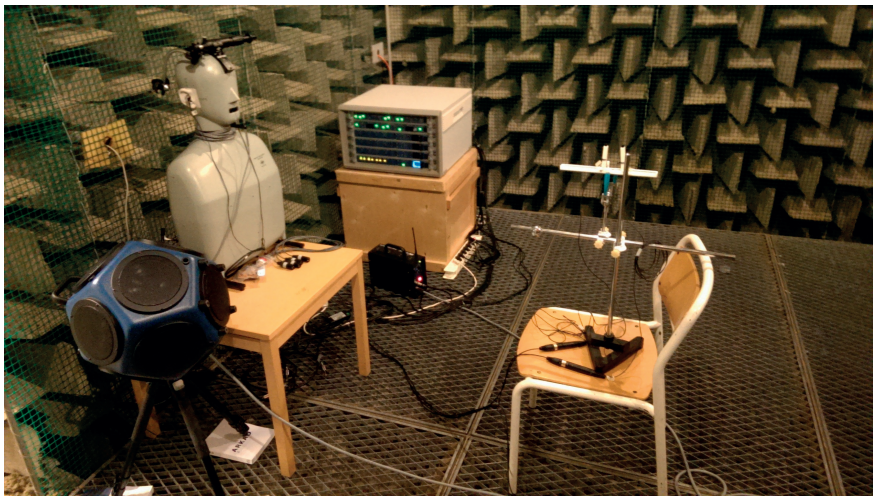

Figure 4.36: The set-up with two sources used to estimate the mixing matrices. The Norsonic 270H was placed to the right of the HATS.

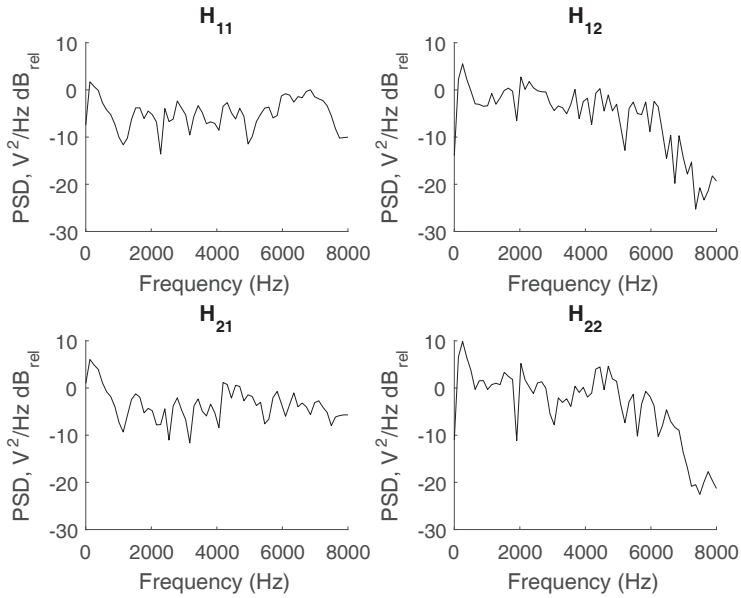

Figure 4.37: The frequency response for the array with a parabolic reflector with a diameter of 10 cm. The noise source is placed directly to the right of the HATS.

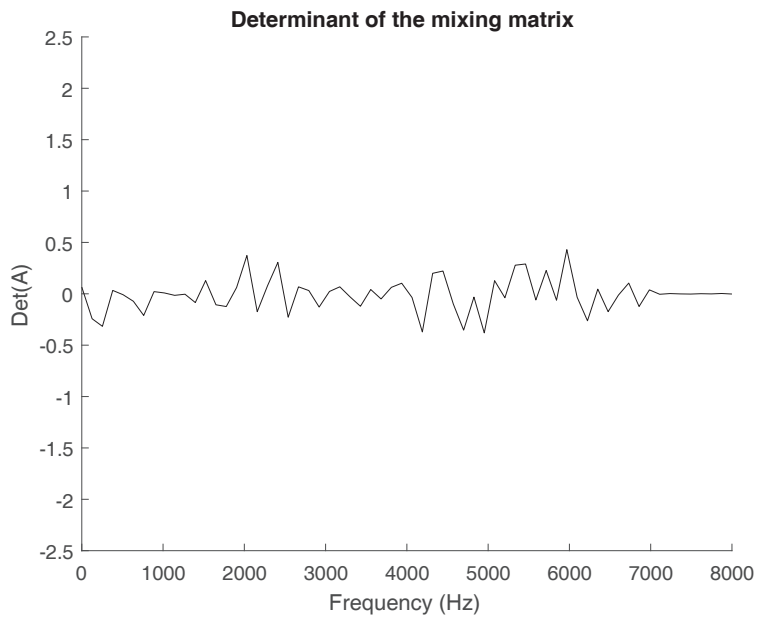

Figure 4.38: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 10 cm. The noise source is placed directly to the right of the HATS.

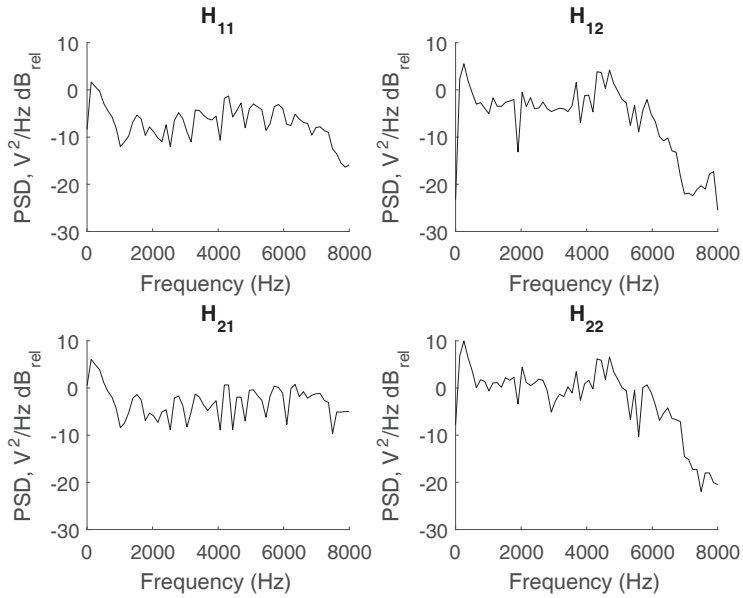

Figure 4.39: The frequency response for the array with a parabolic reflector with a diameter of 5 cm. The noise source is placed directly to the right of the HATS.

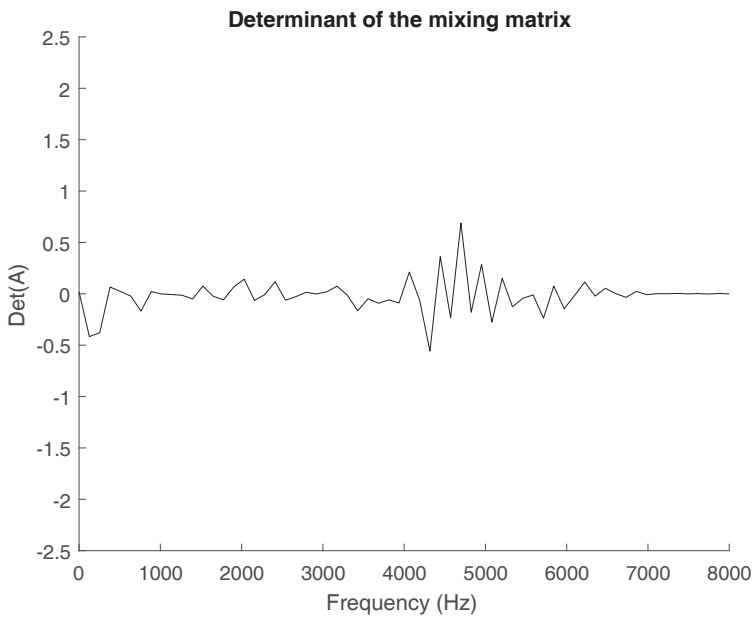

Figure 4.40: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 5 cm. The noise source is placed directly to the right of the HATS.

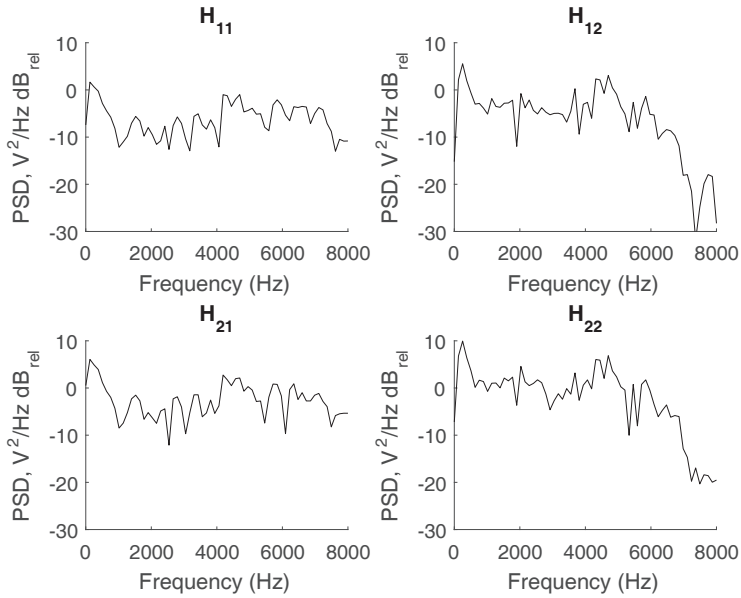

Figure 4.41: The frequency response for the array with a parabolic reflector with a diameter of 2.5 cm. The noise source is placed directly to the right of the HATS.

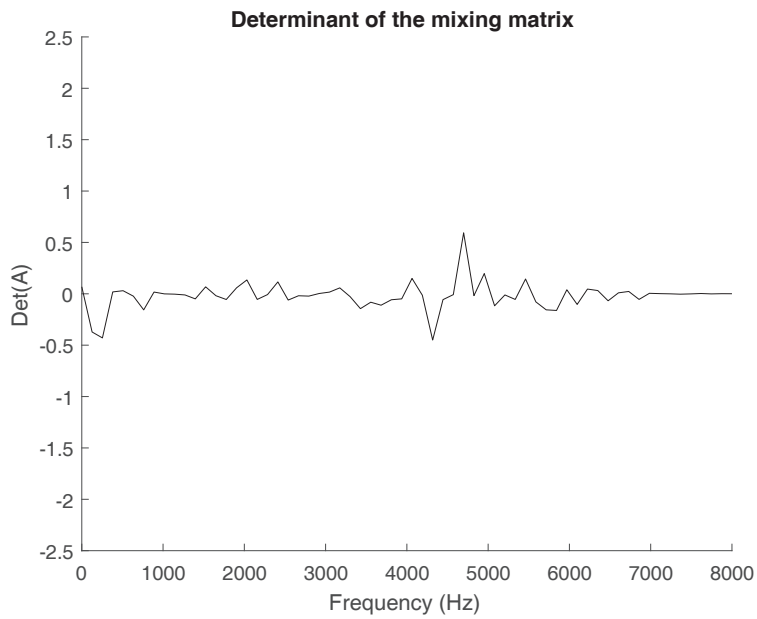

Figure 4.42: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 2.5 cm. The noise source is placed directly to the right of the HATS.

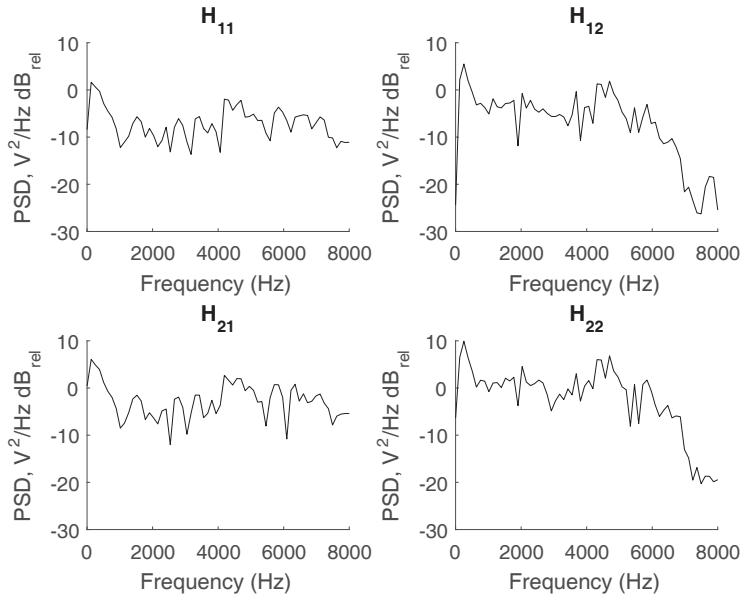

Figure 4.43: The frequency response for the array without a reflector. The noise source is placed directly to the right of the HATS.

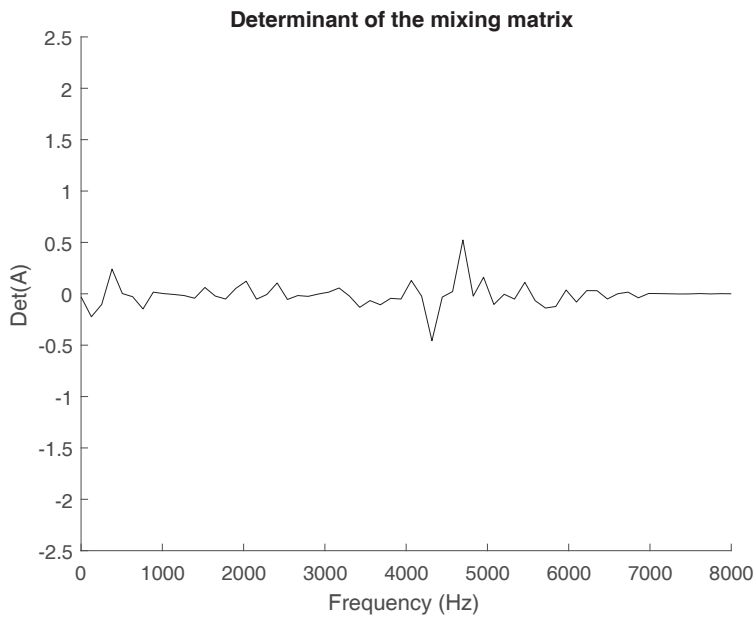

Figure 4.44: The mixing matrix determinant for the array without a reflector. The noise source is placed directly to the right of the HATS.

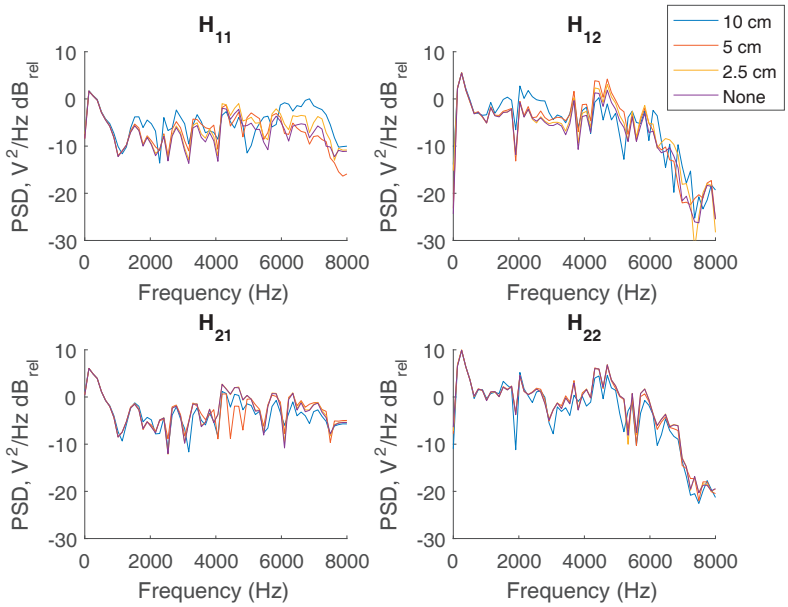

Figure 4.45: The frequency responses for two sources and two microphones when the noise source is placed directly to the right of the HATS.

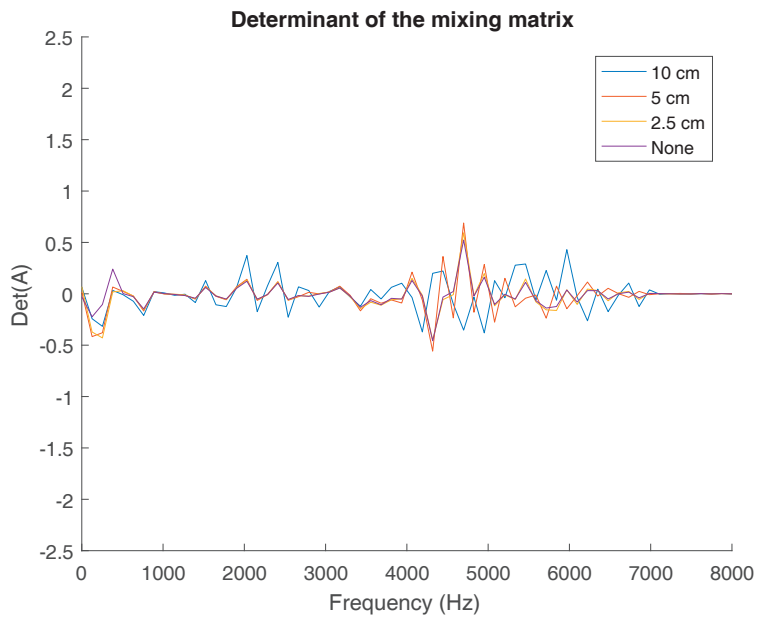

Figure 4.46: The determinants of the mixing matrices in linear scale when the noise source is placed directly to the right of the HATS.

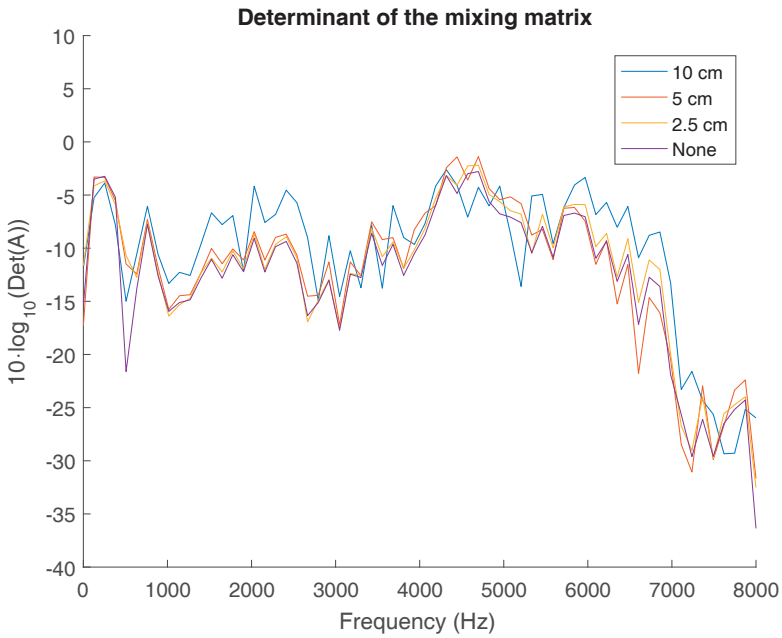

Figure 4.47: The determinants of the mixing matrices in logarithmic scale when the noise source is placed directly to the right of the HATS.

4.3.4 New mount on the 5 cm reflector

A new microphone holder as shown in Figure 4.48 was made for the 5 cm reflector. The idea is that this will improve the mixing channel in two ways. First, that the opening of the microphone is now aimed directly towards the reflected waves which should increase the gain for these. Secondly, mechanical vibrations induced into the reflector can be transferred to the microphone. A similar mount which aimed the microphone towards the reflector was made for the 10 cm reflector. The measurements showed that this new mount made very little difference to the frequency response. The measurements were made using the same configuration as in set-up 1, Figure 4.12.

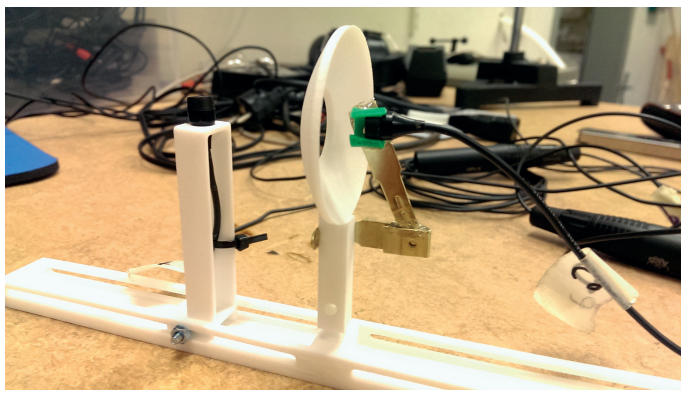

Figure 4.48: A new microphone mount was made for the 5 cm reflector attaching the microphone directly to the reflector and aiming it towards the reflector.

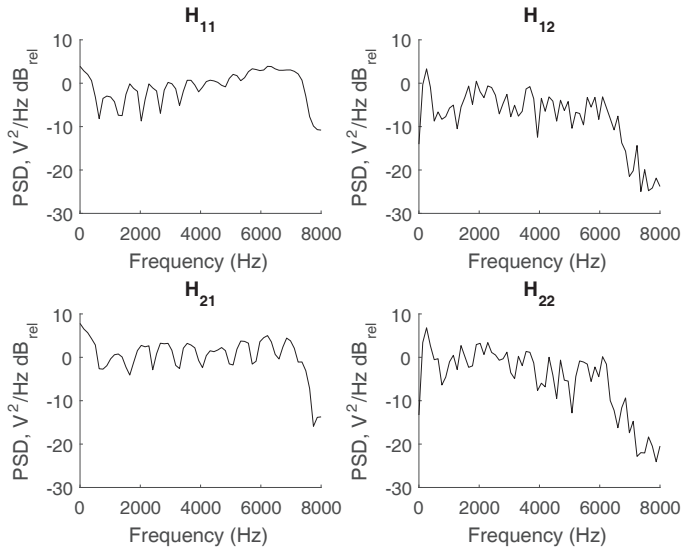

Figure 4.49: The frequency response for the array with a parabolic reflector with a diameter of 5 cm, using the new microphone mount. The noise source is placed behind the array.

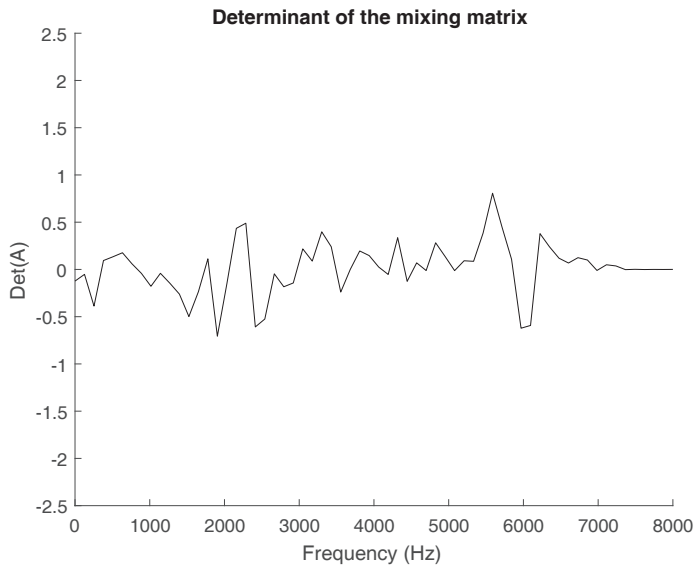

Figure 4.50: The mixing matrix determinant for the array with a parabolic reflector with a diameter of 5 cm, using the new microphone mount. The noise source is placed behind the array.

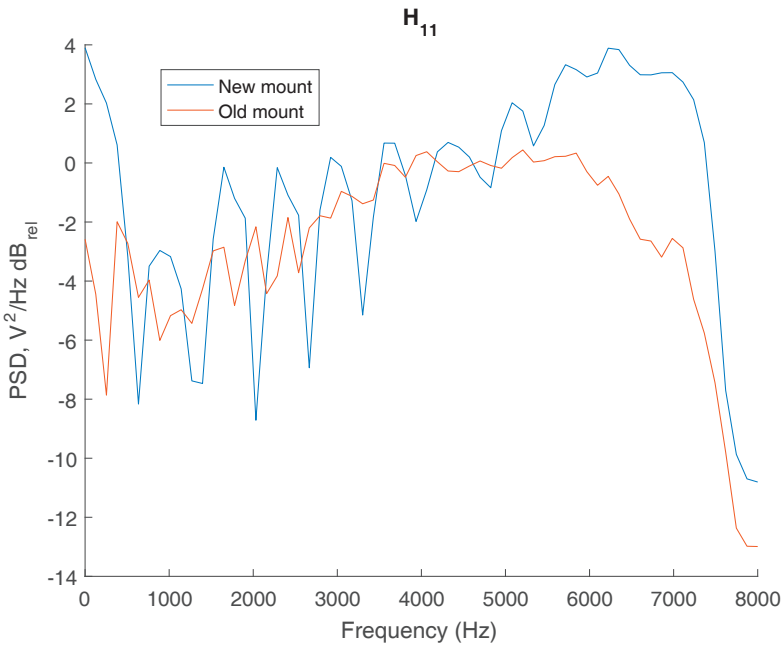

Figure 4.51: The frequency response of the microphone mounted in the new and the old mount.

4.4 Speech evaluation using PESQ

4.4.1 Measurements

Measurements were made using all three reflectors and without a reflector. White noise was played by both the HATS speaker and the Norsonic 270H, this was recorded to be used for channel estimation. The array was set up so that the HATS was about 0.5 m in front of the reflector and the Norsonic 270H 1 m behind the reflector.

4.4.2 Un-mixing of speech passed through the estimated channels

The mixing channels were estimated using the recorded white noise in the H_1 estimator, and then inversely Fourier transformed into 64-tap FIR filters. The inverses of these filters are then calculated using the Wiener solution for the inverting channels. A 1000-tap un-mixing filter was calculated.

A speech signal and a noise signal was mixed using the estimated filters and then unmixed using the calculated inverse filters. This unmixed speech signal was then evaluated using PESQ. Recordings of three different female and three different male speakers uttering a number of different English sentences was used. A 10 second speech signal was generated for each speaker by concatenating the recorded sentences. The power of the speech signals vary depending on the speaker and they were all mixed with noise of the same power, thus resulting in different signal-to-noise ratios (SNR).

4.4.3 PESQ results

The results when running PESQ on the mixed and un-mixed speech signal can be seen in Tables 4.1 and 4.2 respectively. The determinants of the matrices used for the mixing can be seen in Table 4.3.

Speaker	10 cm	5 cm	2.5 cm	No reflector	5 cm, new mount	SNR
Female 1	1.899	1.658	1.579	1.512	1.613	-5.04 dB
Female 2	2.060	1.707	1.666	1.520	2.242	3.53 dB
Female 3	1.781	1.462	1.386	1.299	1.438	-5.24 dB
Male 1	2.129	1.851	1.797	1.710	2.411	3.65 dB
Male 2	2.274	1.981	1.980	1.854	2.287	1.04 dB
Male 3	2.028	1.710	1.710	1.601	2.390	4.35 dB

Table 4.1: The PESQ of the mixed and untreated signals.

Speaker	10 cm	5 cm	2.5 cm	No reflector	5 cm, new mount	SNR
Female 1	1.701	1.530	1.578	1.598	1.935	-5.04 dB
Female 2	1.982	1.829	1.913	1.895	2.217	3.53 dB
Female 3	1.495	1.268	1.290	1.305	1.790	-5.24 dB
Male 1	2.330	2.171	2.162	2.123	2.619	3.65 dB
Male 2	2.237	2.124	2.204	2.182	2.593	1.04 dB
Male 3	2.374	2.201	2.252	2.265	2.549	4.35 dB

Table 4.2: Results from the PESQ-evaluation with a 64-tap mixing filter and a 1000-tap inverse filter. SNR are for the signals sent in to the mixing channels.

	10 cm	5 cm	2.5 cm	No reflector	5 cm, new mount
Median det	0.3160	0.3065	0.2526	0.2154	0.2574
Mean det	0.4457	0.3863	0.3390	0.2969	0.3446

Table 4.3: Median and mean determinants of the mixing matrices used for mixing the signals for the PESQ.

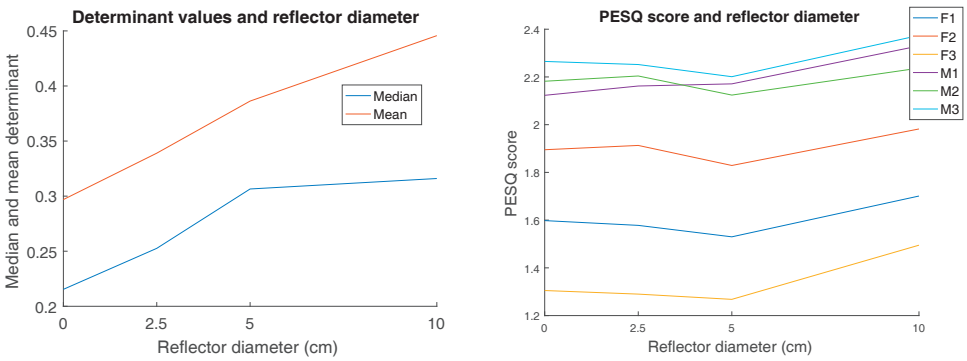

Figure 4.52: The mixing matrix determinant and PESQ scores plotted against the reflector diameter.

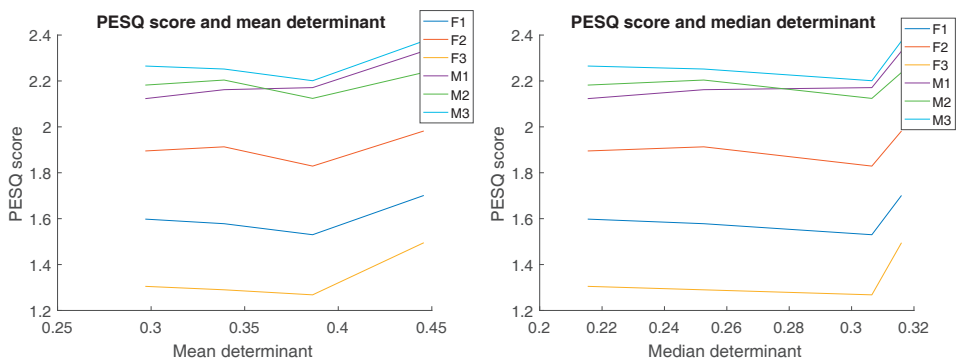

Figure 4.53: The PESQ scores plotted against the mean and median determinants of the mixing matrices.

4.5 EarIn device

The EarIn earplugs have three microphones in total, two are mounted in the left earplug, one facing forward and one facing into the ear canal. The third microphone is mounted in the right earplug facing outwards. For the measurements with two sound sources the left and right outer microphones were used. The left one is set as microphone 1 and the right one as microphone 2.

First the frequency repose of the microphones mounted in the earplugs was measured. This was made by hanging the earplug by its cord in a stand and turning the earplug so that the opening for the microphone was facing the mouth of the HATS. The distance between the mouth of the HATS and the earplug was 11 cm. These measurements were made for both the left and right outer microphones on the earplugs, playing back 10 seconds of flat-spectrum uncorrelated noise in the HATS mouth, recording with a sample rate of 48 kHz. The gain levels were set to 20 dB both for the mouth reference microphone and for the two earplug microphones. In the figures, the spectrum is plotted from 0 Hz to 8 kHz, to make it easier to compare these spectrum to the ones for the reflector array, which were measured using a sample rate of 16 kHz.

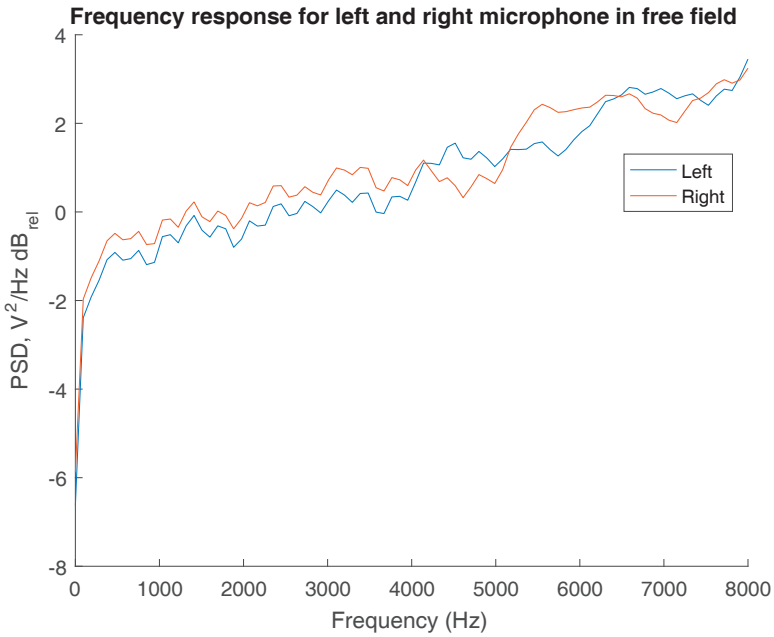

Figure 4.54: The frequency response of the earplug microphones in free field.

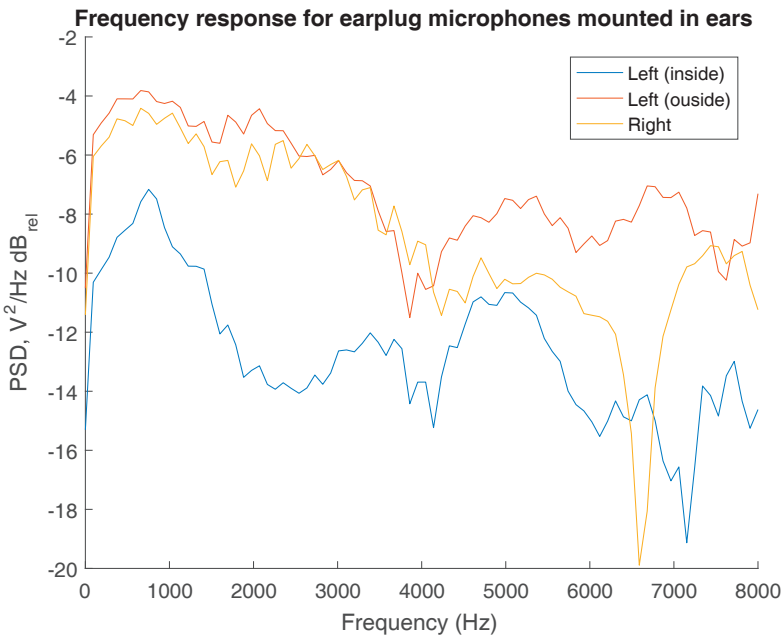

Figure 4.55: The frequency response of the earplug microphones mounted in the ears of the HATS.

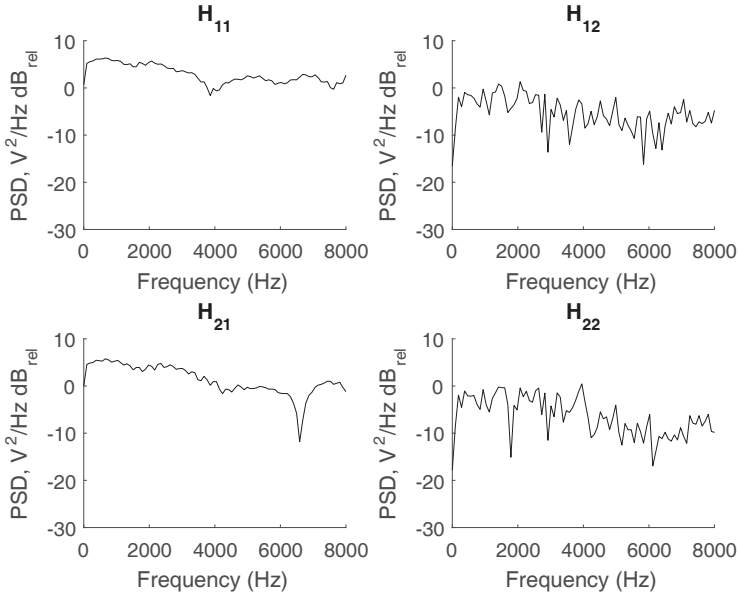

Figure 4.56: The frequency response of the earplug microphones for two sound sources. The mouth of the HATS and a speaker placed in front of the HATS.

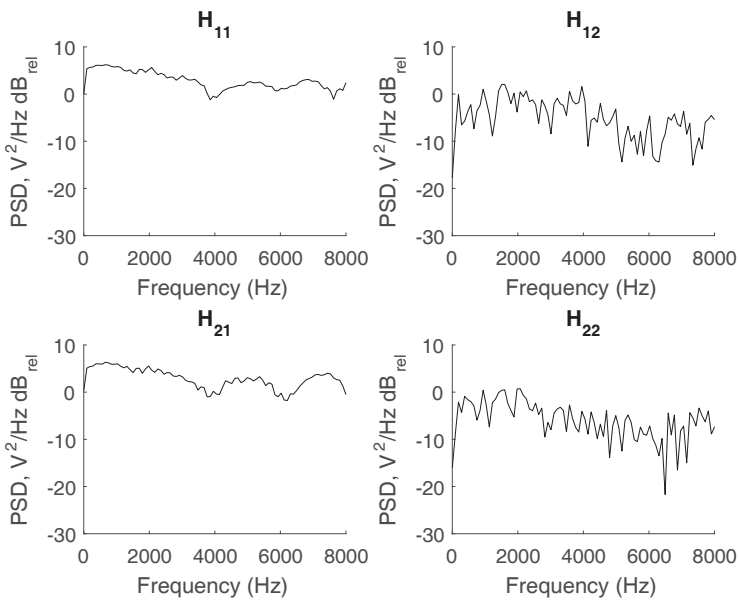

Figure 4.57: The frequency response of the earplug microphones for two sound sources. The mouth of the HATS and a speaker placed 45° to the left of the the HATS.

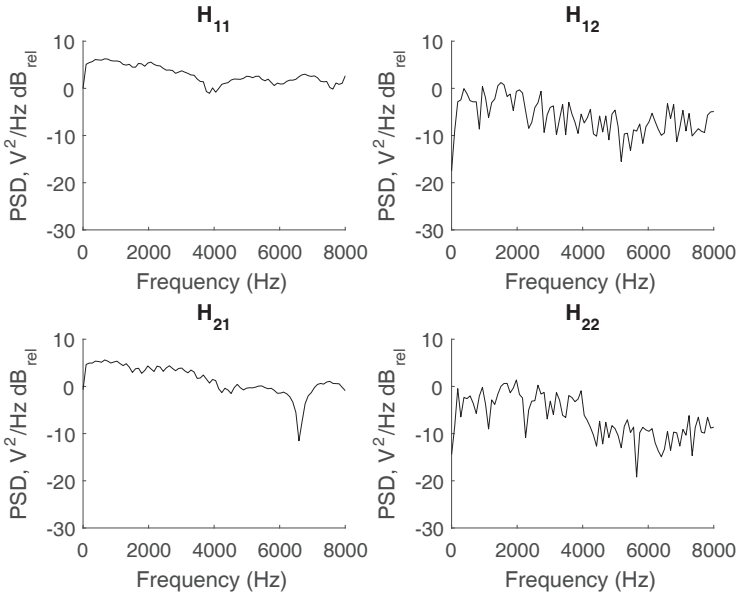

Figure 4.58: The frequency response of the earplug microphones for two sound sources. The mouth of the HATS and a speaker placed 45° to the right of the the HATS.

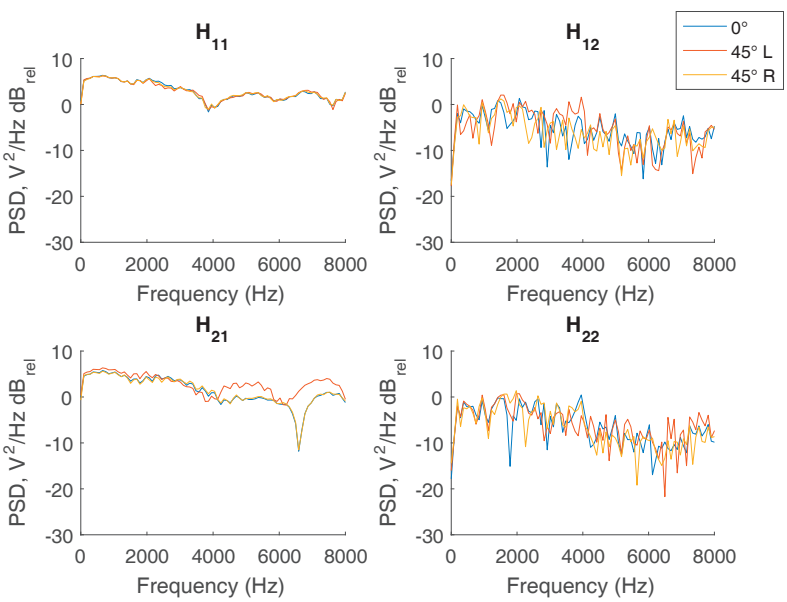

Figure 4.59: The frequency responses for two sources and two microphones. The second source was placed in front of, and 45° to the left and right of the HATS.

Chapter 5
Discussion

5.1 Measurement methodology

The shotgun microphone is commonly used for film and TV recordings when a directional microphone is needed. For this thesis a parabolic reflector was chosen over a shotgun microphone since the reflector offers more freedom of design and most important, the reflector can be made more compact than an interference tube. A shotgun microphone can also be sensitive to sound coming from behind, unlike the reflector.

An improvement that could have been made to the measurements made to determine the mixing channels would have been to use a reference microphone placed close to the Norsonic 270H and use the signal from this microphone as the input when estimating the channels. Now the signal output by Matlab, i.e. a flat spectrum noise signal between 0 and 8000 Hz, was used as the second input for the channel estimation. This means that the channels $H_{12}(f)$ and $H_{22}(f)$ contain both the filtering that occurs in the Norsonic 270H and in that which is caused by the structure and the room. No speaker can output a perfect representation of its input signal, and in Figure 4.3 it can clearly be seen that the Norsonic 270H changes the input signal. We are really only interested in what happens from when the sound leaves the speaker until it is picked up by one of the AKG microphones.

The cavity in the HATS filters the output signal. This filtering was bypassed by using the signal from a reference microphone mounted by the mouth of the HATS as the input for the estimations. $H_{11}(f)$ and $H_{21}(f)$ should thus represent the transfer from when the sound leaves the mouth of the HATS until it reaches one of the AKG microphones including the filtering of the reference microphone. $H_{11}(f)$ and $H_{21}(f)$ will also contain the filtering performed by the AKG microphones. But as can be seen in Figure 4.2 they should have a reasonably flat frequency response at least up to 5000 Hz and then increase slightly. Looking at the measured frequency response for the AKG microphone, in Figure 4.4, the gain increases slightly

with the frequency of the sound. The x-axis here is in linear scale and the one provided by the manufacturer is in logarithmic scale, which has to be taken into consideration when comparing the graphs. The frequency response of the AKG C417 was considered to be sufficiently flat and it was not compensated for in the succeeding measurements.

Care was taken to measure out so that microphone was placed in the focal point of the reflectors. This was made using a ruler with millimetre grading, so it should be accurate, however there is always the possibility of errors such that the microphone is placed out of focus. If the reflectors were slightly warped during manufacture the focal point would be smudged out, however they showed no signs of being defective, so this is assumed to not be a problem.

5.2 Polar patterns

Looking at the measured directional frequency responses, Figures 4.6, 4.8 and 4.10 and the polar plots, Figures 4.7, 4.9 and 4.11, it can be seen that the 10 cm reflector has directional properties, especially for higher frequencies. The local minimum at 60° in Figure 4.7 is probably caused by that the reflector is blocking the direct path to the microphone at this angle. The polar patterns for the smaller reflectors show that they are almost omni-directional. The measurements for the polar plots were quite crude and made in coarse increments. A proper controllable rotation table wasn't available at the time, so the array had to be rotated manually, entering the anechoic room to move it between each measurement.

Even though the measurements are crude and inexact, it can be concluded that the smaller reflectors are nearly omnidirectional, otherwise some different gains for different angles should be seen, like for the 10 cm reflector. The measurements of the polar patterns could be improved by taking angular measurements in smaller increments.

5.3 Frequency responses and determinants

The frequency responses calculated using the ray-tracing model differ from the measured frequency responses. The zeros that show up in the calculated frequency responses are not as clear in the measured frequency responses.

Looking at Figure 3.1, the transfer function for the 10 cm reflector has a zero around 1500 Hz that isn't visible in Figures 4.13, 4.25 and 4.37, the measured frequency responses. This is probably caused by that the real reflector reflects very little at 1500 Hz, so there is not much interference in this frequency range. In the measured frequency response there is a dip at

about 4500 Hz, this could very likely be an overtone to the zero at 1500 Hz. Which agrees to the fact that the amount of reflected content increases and the amount of diffracted content decreases as the frequency increases.

The ray model is generally considered accurate when the reflecting structure is much larger than the wavelength of the reflected wave. In this case the wavelengths are of similar size and even larger than the structure, so the ray-model may be inaccurate. The efficiency factor of 0.7 that was used for the simulations was probably too large, a bigger part of the energy got absorbed by the structure in reality.

An interesting result can be seen in Figures 4.21, 4.33 and 4.45, at the higher frequencies where $H_{11}(f)$ increases, $H_{21}(f)$ decreases, this means that more of the incoming sound is reflected into microphone 1 rather than being diffracted around the reflector into microphone 2.

Looking at Figures 4.22, 4.23, 4.34, 4.35, 4.46 and 4.47, the determinants were the biggest when the noise source was placed behind the array, and smallest when the noise source was placed just to the side of the HATS, and having the second source 90° to the right was somewhere in between. Having the two sources close to one and other is the hardest situation to separate. Both because some of the sound from the second source will be reflected in the parabolic reflector, and there is very little phase difference between the two sources in this case. This agrees with the fact that the determinant is smaller in this case.

The simulations generally showed that the determinants of the mixing matrices were growing faster and more monotonically than for the measured values. This is caused partly by that the simulated reflector was set to be too efficient, and partly by that the sources and sensors are assumed as points, which means that there is virtually no variation in phase of a signal that arrives at the sensor. That is, the additional distance the reflected wave travels in the simulations is always $2L$ where in reality the distance is $2L \pm$ the diameter of the microphone diaphragm. In the measurements there is also the drop-off above 7500 Hz caused by the filtering applied by the recording software, which is not present in the simulations.

Looking at the Figures for the 5 cm and 2.5 cm reflectors and the array without a reflector, 4.15, 4.16, 4.17, 4.18, 4.19, 4.20, for set-up 1, 4.27, 4.28, 4.29, 4.30, 4.31, 4.32 for set-up 2 and 4.39, 4.40, 4.41, 4.42, 4.43, 4.44 for set-up 3. It can be seen that the smaller reflectors doesn't affect the frequency responses and determinant values considerably. This is caused by that they are too small to reflect frequencies below 8000 Hz, and waves diffract around the reflectors instead of being reflected. Looking at Figures 4.14, 4.26 and 4.38 it can be seen that determinant values were largest when using the 10 cm reflector. The determinant was the largest when the noise was placed behind the array, which agrees with that this set-up gave

the best conditioned problem, not counting the 5 cm reflector with the new mount.

As can be seen in Figures 4.49 and 4.51, modifying the mount so that the microphone was aimed towards the reflector and attached to the reflector made a considerable difference in the frequency response. The gain for the 5 cm reflector increases with frequency like the response for the 10 cm reflector when using the new mount. The reflected waves now have a more direct path to the microphones diaphragm. The diaphragm is placed directly in the focal point and is stretched out vertically compared to being stretched out horizontally which means that the diaphragm will be placed in the focal point $\pm$ the diameter of the diaphragm when using the old mount. It is possible that the increased gain between 0 and 500 Hz is caused by mechanical vibrations being transferred to the microphone. There are some sharp zeros in the new frequency response which may have a negative effect on the recorded signal.

5.4 PESQ

Figure 4.52 shows that the median and mean determinant values are positively correlated to the reflector diameter. In Figure 4.53 it can be seen that the PESQ-scores are low for all speakers for the 5 cm reflector, and the highest for the 10 cm reflector. It seems like the channels for the 5 cm reflector was particularly hard to un-mix.

Even though the determinant increases with reflector diameter, the determinant is not the only factor that affects how easily the sources are separated, and it looks like the reflector has to be bigger than 5 cm to increase the PESQ-score. For the 5 cm reflector with the new mount determinant was actually smaller, see Figure 4.50, but the separation was better. Listening to the un-mixed signals, the ones for no reflector, 2.5 cm and 5 cm reflectors sounds roughly similar, but in the one for the 10 cm reflector the noise is more attenuated. This shows that for a large enough reflector it is easier to separate the recorded signals, but if the reflector is smaller than 5 cm in diameter is makes very little difference. However there is still quite a lot of noise even in the separated mixture for the 10 cm reflector. The mixing filter was just 64 taps, and the inverse was 1000 taps, but it could not un-mix it very well, showing just how hard this a convoluted mixture is to separate.

For the PESQ evaluation, the mixing channels were first identified, and the inverting channels were calculated. Then speech and noise was run through first the mixing channels and then through the inverting channels. This method was chosen instead of recording speech and noise in the ane-

choic room, and then trying to un-mix these using the calculated inverses. The main reason behind this is that there are a lot more factors that come into play when trying to separate a real recording.

The calculated inverse is the inverse of the estimated channel, thus mixing with the estimated channel is the best possible condition. Un-mixing real recorded signals requires that the estimated channels are a very good representation of the real situation in the anechoic room. It also requires that the relations between the gain levels in $H_{11}(f)$, $H_{12}(f)$, $H_{21}(f)$ and $H_{22}(f)$ represent the real situation. Otherwise the un-mixing filter would be either too aggressive, distorting the speech signal, or not strong enough, removing too little of the noise.

There seems to be some correlation between the PESQ score and mixing matrix determinant for reflectors bigger than 5 cm. However, to confirm this hypothesis, reflectors between 5 and 10 cm would have to be tried, as well as reflectors bigger than 10 cm.

Looking at Tables 4.1 and 4.2 it can be seen that for the speakers Female 1 and Female 3 the PESQ score is actually lower for the un-mixed signals regardless of the reflector size. This shows that when the SNR is very low, the un-mixing filters actually distort the signal rather than improve it. In Table 4.1 it can also be seen that the PESQ scores are increased just by putting a reflector in the array, which seems plausible since the reflector amplifies the speech signal.

Changing the mount for the microphone had a positive effect on the PESQ scores. Both the PESQ scores for the mixed and unmixed signals are better for the new mount comparing those scores to the ones for the old mount with the 5 cm reflector. They are even slightly better than the ones for the 10 cm reflector with the old mount. However it has to be taken into consideration that the array was taken apart between the measurements and set up again on a different day. This may have affected the results, and an absolute conclusion cannot be drawn without repeating the experiments using the old mount under the exact same conditions. Based on the results it can be seen that the 5 cm reflector has a bigger effect on wavelengths in the voice band than was previously stated when using the new microphone mount. Even a reflector with a diameter of 5 cm can have a positive impact on the un-mixing of the signals, but the direction in which the microphone is aimed is important for smaller reflectors. The sharp zeros, as seen in Figure 4.51 that was added when the mount was changed doesn't seem to have affected the PESQ score negatively.

5.5 EarIn device

First of all, when the EarIn devices are placed in free field the gain of the microphones increase with increasing frequency. This might be caused by that the microphones are mounted in ports with a small diameter of the opening. Analogous to the fact that how much sound an object reflect is related to the wavelength of the sound in proportion to the size of the object, it is easier for high frequency sound to enter an opening of a small diameter, explaining the slope in the frequency response.

The channel identification on the EarIn earplugs showed that the channels change when comparing hanging the microphones in free-field and mounting the in the ears of the HATS, compare Figures 4.54 and 4.55. This is probably caused by reflections on the HATS ear and head, and the dampening caused by directing one of the microphones into the ear canal. The microphone facing into the ear canal is attenuated compared to the outwards facing microphones. Both the left microphones have a minimum at 4000 Hz, the inwards facing microphone has an additional minimum at 7000 Hz. The right microphone has a sharp minimum just above 6500 Hz.

Figures 4.56 - 4.59 show that $H_{12}(f)$ and $H_{21}(f)$ are similar regardless of the direction of the noise, which they should be since they represent the transfer from the mouth to the ear microphones, which should be unchanged regardless of the noise direction. The sharp minimum in $H_{21}(f)$ is however not present when the noise is coming from 45° to the left. This could be caused that the position of earpiece in the ear was slightly changed when rotating the HATS, thus changing how the sound reflects in the ear before reaching the microphone. This shows that using estimated transfer functions from the mouth to the ears to equalize the sound picked up by the ear microphone, has a few problems, since the transfer function is dependant on how the earpiece is placed in the ear, and would thus change every time the user takes out and puts back the earpieces in the ears. The transfer functions from the second source, $H_{12}(f)$ and $H_{22}(f)$, are slightly different depending on the direction of the noise source.

On a real person, a microphone facing into the ear canal will mainly pick up sound carried by the bones and tissue of the person, and not so much of the sound transferred by the air on the outside. The signal to the inner microphone will be low pass filtered, since bone and tissue transfer low frequency sounds better. Using a physical structure to amplify the high frequency sounds coming from the direction of the mouth of the speaker, and combing this with the low-pass filtered signal from inside the ear canal might be a way to improve the sound quality of the recorded and processed speech. However the results of this thesis shows that the structure would have to be at least 5 cm in diameter, which is far too big to be mounted in

a small earpiece. Therefore it can be concluded that the use of reflective is not feasible for this type of application.

5.6 Conclusions

To conclude it can be said that small structures generally have very little effect on sound in the voice frequency range. The sound diffract around small objects rather than being reflected. Somewhere around where the diameter of the structure is about half the wavelength of the sound it starts to reflect against the structure. If one wants to use a reflector for the purpose of separating speech signals, it needs to be at least somewhere between 5 and 10 cm in diameter and as the diameter increases the separation should improve, at least up to some limit. The 2.5 cm reflector had very little effect on the source separation.

There is definitely some potential in using reflective structured for signal separation purposes. The separation results improved for the biggest reflector, and stayed the same or in the case of the 5 cm reflector were slightly worse than when not using a reflector. This shows that parabolic reflectors can be used to improve the results of signal separation, but they have to be so big that they add a sufficient amount of amplification and reflect waves in the frequency range of the signal which one is interested of separating. The properties of the microphone also plays a part for the separation, and in the case of the AKG C417, the separation was improved when it was attached to and aimed towards the reflector.

There is a lower limit in the size where the reflector is effective, and this limit is too big to be used in small portable communication devices and hearing aids. However the reflector could be used in applications where there are less restrictions on size. For example in a voice-control system for a car, a reflector aimed at the driver could be discretely integrated into the interior design of the car, or for table-top conference room phones where a motorized reflector could be aimed at the person who is currently talking.

Chapter 6
Further research

This thesis has confirmed that there are lower limits how small a reflector can be to reflect sound in the voice frequency band, and the required reflector diameter is simply too large to be used in hearing aids or small communication devices.

There are a number of areas where the results of this thesis can be used for further research. The same concept can be expanded upon by for example finding ways to improve the reflective properties of the structure, or find ways to better transfer the mechanical vibrations induced by the sound into the structure to a microphone, or using an additional piezoelectric microphone to pick up these vibrations.

Using the overtones in a sound signal might also be a possible development for the use of physical structures. Since the overtones have a higher frequency they are reflected by smaller structures, so this could be a way to decrease the size of reflector. The overtones have far lower amplitude than the main part of the signal. Developing a microphone that is only sensitive to very high frequencies, or a way to only record overtones could be of use here.

Ultrasound used for medical imaging has a frequency from 20 kHz up to several GHz. These frequencies would be reflected by a small structure, so it is possible that some kind of small physical structure could be used in ultrasonic probes to improve quality of the measurements.

Acoustic meta-materials is a relatively new area of research, but structures of these materials can supposedly reflect wavelengths that are far larger than the structure. This might be used to make very small reflectors that would fit in hearing aids or communication devices [37][38].

Bibliography

[1] Elko, Gary W., *A Simple Adaptive First-Order Differential Microphone*, Acoustics and Speech Research Department, Bell Labs, Lucent Technologies, Murray Hill, NJ

[2] Takiguchi, T.; Takashima, R.; Ariki, Y., *Active Microphone with Parabolic Reflection Board for Estimation of Sound Source Direction*, Hands-Free Speech Communication and Microphone Arrays, 2008.

[3] Pu, Chiang-Jung, *A Neuromorphic Microphone for Sound Source Localization* Dissertation presented to the Graduate School of the University of Florida, 1998

[4] Ono, N.; Zaitsu, Y.; Nomiyama, T.; Kimachi, A.; Ando, S., *Biomimicry Sound Source Localization with "Fishbone"*, IEEJ Transactions on Sensors and Micromachines 121(6):313-319, January 2001

[5] Batteau, Dwight, W., *The Role of the Pinna in Human Localization*, The Royal Society, 1967

[6] Fredriksen, Erling, *Condenser Microphones*, Book chapter in Handbook of Signal Processing in Acoustics, Springer, 2008, p. 1247-1265

[7] Kadis, Jay, *Microphones*. Educational material, Stanford University, Oct 2004.

[8] Olson, Harry F., *On the Collection of Sound in Reverberant Rooms with Special Reference on the Application of the Ribbon Microphone* Proceedings of the Institute of Radio Engineers, 1933.

[9] Olson, Harry F., *Elements of Acoustical Engineering* D. Van Nostrand Company, 1957, p. 246, 264

[10] `http://en.wikipedia.org/wiki/Carbon_microphone`, visited 2016-05-12.

[11] http://en.wikipedia.org/wiki/Microphone#Piezoelectric_microphone, visited 2016-05-12.

[12] *A New Kind of Laser Microphone Using High Sensitivity Pulsed Laser Vibrometer* Chen-Chia Wang; Sudhir Trivedi ; Feng Jin ; V. Swaminathan, 2008 IEEE International Conference on Mechatronics and Automation

[13] Zhe Wang; Quanbo Zou; Qinglin Song; Jifang Tao, *The era of silicon MEMS microphone and look beyond*, 2015 Transducers - 2015 18th International Conference on Solid-State Sensors, Actuators and Microsystems (TRANSDUCERS)

[14] Burroughs, L.R., *Microphone Facts for the Operating Engineer* Electro-Voice Inc., 1960

[15] *Blind Signal Separation using Overcomplete Subband Representation*, Grbić, N.; Tao, X. J.; Nordholm, S.; Claesson, I., IEEE Transactions on Speech and Audio Processing, 2001.

[16] http://upload.wikimedia.org/wikipedia/commons/4/44/Interference_Tube.jpg

[17] Hyvärinen, A.; Karhunen, J.; Oja, E., *Independent Component Analysis*, John Wiley & Sons, Inc., 2001, p. 356.

[18] Holst, Anders; Ufnarovski, Victor, *Matrix Theory*, KFS AB, 2013, p.57

[19] C., Capdessus; A.K., Nandi; and N., Bouguerriou, *A New Source Extraction Algorithm for Cyclostationary Sources*, Chapter in: Independent Component Analysis and Signal Separation, 7th International Conference, ICA 2007, London, UK, September 9-12, 2007. Proceedings, p. 145-151, Springer Berlin Heidelberg, 2007

[20] Bendat, Julius S.; Piersol, Allan G., *Engineering Applications of Correlation and Spectral Analysis*, Second Edition, John Wiley & Sons, Inc., 1993

[21] Welch, P., *The use of fast Fourier transform for the estimation of power spectra: A method based on time averaging over short, modified periodograms*, IEEE Transactions on Audio and Electroacoustics, 2003

[22] Lindgren, G.; Rootzén, H.; Sandsten, M., *Stationary Stochastic Processes for Scientists and Engineers*, CRC Press, 2014, p. 251-254

[23] http://cdn1.rode.com/ntg2_datasheet.pdf, visited 2016-05-16.

[24] http://sv-se.sennheiser.com/global-downloads/file/802/ME_67_GB.pdf, visited 2016-05-16.

[25] Everest, F.; Pohlmann, Ken C., *Master Handbook of Acoustics* Fifth Edition, McGraw Hill, 2009, p.99, 107

[26] Chen, Zhixin; Maher, Robert C., *Parabolic Dish Microphone System*, Montana State University

[27] Skålevik, M., *Low Frequency Limits of Reflector Arrays*, akuTEK, 2007

[28] Talbot-Smith, Michael, *Audio Engineer's Reference Book*, Second Edition, Focal Press, 1999, p. 32.

[29] Nilu Mishra; Archana Charkhawala, *Human recognition through RFID: A distinct application of speech processing*, 2011 3rd International Conference on Electronics Computer Technology (ICECT).

[30] http://en.wikipedia.org/wiki/Voice_frequency, visited 2016-05-12.

[31] http://www.bnoack.com/index.html?http&&&www.bnoack.com/audio/speech-level.html, visited 2016-05-12.

[32] Kuttruff, Heinrich, *Room Acoustics*, Fifth Edition, Spon Press, 2009, p. 58.

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

[34] http://cloud.akg.com/8256/c417_manual.pdf, p.23., visited 2016-05-11.

[35] http://www.bksv.ro/produse/Head%20and%20Torso%20Simulators%20-%20Types%204128-C%20and%204128-D.pdf, p.3., visited 2016-05-11

[36] http://www.campbell-associates.co.uk/userguides/Nor250-270-280userEn.pdf, p.5., visited 2016-05-12.

[37] https://www.newscientist.com/article/dn13156-acoustic-superlens-could-mean-finer-ultrasound-scans, visited 2016-05-27.

[38] Zhang, Shu; Yin, Leilei; Fang, Nicholas, *Focusing Ultrasound with Acoustic Metamaterial Network* Phys. Rev. Lett. 102, 194301, 2009

LUND
UNIVERSITY

Series of Master's theses
Department of Electrical and Information Technology
LU/LTH-EIT 2016-527
http://www.eit.lth.se