Acoustic Room Correction for Speaker Systems Using Signal Processing Techniques

Niklas Carlsson
dat13nca@student.lu.se
Hanne Heingård
geo13hhe@student.lu.se

Department of Electrical and Information Technology
Lund University

Supervisor:
Mikael Swartling, Lund University
Danny Smith, Axis Communications
Arvid Nihlgård Lindell, Axis Communications

Examiner: Nedelko Grbic

June 9, 2018

Printed in Sweden
Tryckeriet i E-huset, Lund

Abstract

Loudspeakers are used in many different applications and play a vital role in various types of entertainment and communication. To optimize the listening experience, there is a need for an adjustment process to compensate for any effects the environment might have on the loudspeaker's performance. This thesis presents an adaptive method that utilizes an arbitrary amount of loudspeakers and microphones to calibrate an optimized sound profile for a specific room. The method proposes compensation changes to the loudspeakers combined with the possibility to further adjust the sound according to a specific user profile. Two different approaches are possible in the calibration phase, white noise calibration and tonal signal calibration. Using the proposed method, tests concluded that the sound profile was improved in both approaches. Calibrating using the white noise signal proved to be superior to using the tonal signal. The feasibility of the method is verified through a number of tests in various environments and with different setups by measuring the frequency responses in different points of interest before and after calibration.

Acknowledgements

Special thanks to our supervisors at Axis, Danny Smith and Arvid Nihlgård Lindell, for their support and knowledge with equipment and signal processing. Also thank you to Carl Hansson at Axis for his enthusiastic support at any time needed. Special thank you to Catarina Cedervall for giving us the opportunity to conduct this thesis project at Axis where we have learned something new every day and met new people. Lastly, thank you to all remaining colleagues at the Audio and Door station department at Axis who has helped us through this project and made our time there unforgettable.

Thanks to Martin and Sofie who conducted their thesis project at the same time as we did on Axis and helped us with MATLAB in particular.

From LTH we wish to thank Dr. Mikael Swartling for his help and explanations to any questions we had and for leading us in the right direction, it has been invaluable.

Thank you Dr. Nedelko Grbic for your knowledge and guidance.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

BW Bandwidth

CF Correction Factor

DFT Discrete Fourier Transform

DSP Digital Signal Processor

EQ Equalizer

FFT Fast Fourier Transform

FT Fourier Transform

NTP Network Time Protocol

OS Operative System

PoE Power over Ethernet

PoI Point of Interest

PSD Power Spectrum Estimator

PTP Precision Time Protocol

RMS Root Mean Square

SD Standard Deviation

Chapter 1
Introduction

1.1 Background

When playing sound from loudspeakers many factors affect the listening experience. The room's natural acoustics, such as shape and furniture, as well as the loudspeakers physical limitations and characteristics will have a noticeable impact on sound quality and perception. This results in certain frequencies being attenuated or boosted which affects the listening experience. While some factors can not be practically changed, for instance the physical attributes of the environment, adjustments can be made to the loudspeaker's sound profile to counter the room's acoustic attributes. Approaches to perform these adjustments are generally referred to as acoustic room correction methods which uses digital signal processing in an attempt to accomplish automatic and smart equalization for the loudspeaker.

Axis Communications AB has developed network loudspeaker solutions for several different purposes, such as music streaming and voice announcements. The speakers are a complete PoE driven system integrating a Linux OS, DSP, amplifier and network interface. They are placed either in the ceiling, on a wall or a pillar depending on the model preferred and can cover various locations such as airports, schools, shopping malls etc. As room magnitude differs, as well as quantity of speakers and their placement, Axis has emphasized the need for an automated process of loudspeaker correction to compensate for any disadvantages and disparities in the setup environment. This project made use of a controlled loudspeaker environment at Axis premises as well as an anechoic chamber located at LTH.

1.2 Motivation

The motivation for this thesis is to improve the quality of the generated sound from each individual loudspeaker and a group of speakers, wherever in the room they are physically located. This will be beneficial to the listening experience as well as general sound perception of the whole room. Considering every room's unique dimensions and attributes, it would be of great interest to Axis to implement these features and automate the process as much as possible.

1.3 Objective

The objective with this thesis is to investigate the feasibility of acoustic room correction techniques and how well they are able to compensate for unfavorable speaker locations and adjust settings accordingly. In the end this should result in a mobile phone application to be used by an external installer of the speaker systems. The application is to perform calculations and suggest appropriate quality enhancing operations for each individual loudspeaker as well as speaker groups. It should also take in to account how many loudspeakers the room contains in total, in order to enhance the general room listening experience and the interaction between the loudspeakers. The goal is to be able to automatically, using the phone application, calibrate frequency response and amplifier gain to desired and optimal levels to a set of points of interest in the room.

1.4 Thesis outline

Chapter 1 covers the introduction for the project, its motivation and objectives as well as presenting Axis Communications AB.

Chapter 2 describes the depth of the problem to be solved and any relevant theory regarding room correction and digital signal processing that is later used. It includes information about microphones and mathematical formulas used globally throughout the project.

Chapter 3 covers the hardware and software details for the thesis. Any equipment used is listed and explained as well as the software built during the project and its purposes.

Chapter 4 covers the localization and distance calculation theory, method and results.

Chapter 5 entails the single channel calibration testing. This chapter is one of the main parts of the thesis and covers all theory, measurements and results of the testing done using a described setup.

Chapter 6 has the same outline as previous chapter but describes a multi channel setup consisting of an arbitrary amount of speakers and microphones. Relevant theory and testing results are found in the same chapter.

Chapter 7 summarizes the thesis and its most important objectives and intentions along with a brief analysis and possible future work.

Chapter 2

Introductory theory and problem formulation

2.1 Initial problem formulation

Covering areas of different size with loudspeakers requires sound correction methods in order to ensure an optimized listening experience throughout the room. By placing microphones at points of interest in the room, the frequency response can be measured and compared to a reference curve of how the system's frequency response ought to look like in the specific point. In a clothing retail store, places of interest may for instance be the entrance, the checkout counters and the changing rooms as these are locations where customers continually pass through and therefore requires a better listening experience than less frequently visited spots of the area since the retailer want their customers having a good experience visiting their store. It should also be possible to have different sound levels at these points.

Fig. 2.1 shows an example setup with four loudspeakers where the microphones are placed in two potential points of interest. These points will be locations where sound adjustments will be performed. For future reference in this thesis, a point of interest will be equivalent to a microphone placement.

2.2 Fourier transform

The Fourier transform is a function to decompose a signal into its frequency components [9]. It is defined by

$$F(\omega) = \int_{-\infty}^{\infty} f(t) e^{-it\omega} dt \tag{2.1}$$

where i is the imaginary unit. The FT is a complex-valued function $R \to C$ where the complex part represents the phase offset and the absolute value represents the magnitude of the chosen frequency that was present in the signal. Breaking down a signal into its frequency components using the FT opens up new mathematical possibilities. For example, convolution in the time domain is represented by multiplication after performing the FT. From a signal processing viewpoint, the FT is useful for computing different types of filters where different frequencies should be affected in different ways.

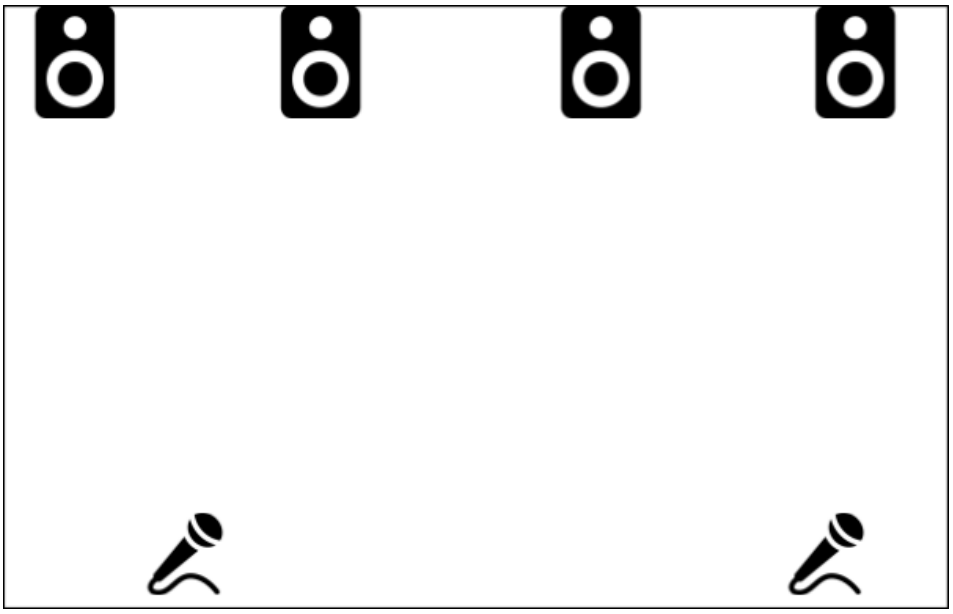


Figure 2.1: Microphone placement in points of interest with four loudspeakers.

2.3 Fast Fourier transform

The fast Fourier transform is effective for calculating the discrete Fourier transform. It is fairly simple to implement the FFT and is a good choice when working with a full frequency spectrum. The FFT computes the DFT which is used to transform a signal from time domain to its frequency domain by using discrete samples instead of a continous signal. Computing the DFT in general has a computational complexity of $O(N^2)$ by using the normal FT, but by instead using the FFT algorithm to perform the calculation this can be reduced to $O(N \log N)$ [29]. Since the FFT is used to compute the DFT, the term FFT will be used for both the FFT and the DFT throughout the thesis. The DFT is defined by

$$X_k = \sum_{n=0}^{N-1} x_n e^{-i2\pi kn/N} \tag{2.2}$$

where $k = 0 \dots N-1$ and where $x_0 \dots x_{N-1}$ are complex numbers. The DFT coefficients in X_k is a finite duration sequence.

2.3.1 Goertzel algorithm

However, when working with single or a few frequencies, the more appropriate choice is to use the Goertzel algorithm for the transformation. The Goertzel algorithm is numerically more efficient than the FFT as long as the number of frequencies M to calculate does not exceed $\log N$, where N is the number of

samples. In that case, the FFT will be a better method in terms of efficiency [26]. The Goertzel algorithm recursively run the equation

$$s_k(n) = 2cos(\frac{2\pi k}{N})s_k(n-1) - s_k(n-2) + x(n) \tag{2.3}$$

where *x(n)* is the input signal. This algorithm will iterate through $n = 0 \dots N$. However, the computation in equation 2.3 is only done once for $n = N$, that is, when a chosen frequency occurs in the extracted data of samples. This leads to the Goertzel algorithm having a favourable time complexity of *O(MN)* where N is the amount of samples and M is the number of frequencies we are interested in compared to the FFT. In the case of only wanting to find a single frequency, the complexity will therefore be reduced to *O(N)*. The output computation in equation 2.3 will generate the magnitude of the chosen frequency in specified sample input. For example, if the sample size is set to 16, the equation will iterate through the input signal 16 samples at a time [27][28]. The objective is to locate the start of the signal of the desired frequency by analyzing its magnitude and comparing it to the magnitude of the input signal which is described by

$$y_k(n) = s(n) - e^{j2\pi k/N}s(n-1) \tag{2.4}$$

2.3.2 Spectral leakage

Spectral leakage defines the term in which new frequencies appear in the output that were not present in the original input when a signal has been passed through a non-linear system. That is, the signal has not been replicated identically to how it was created. This is caused by performing the FFT on non-periodic data. A signal concentrated to a single frequency have been spread out over the entire frequency range, causing spectral leakage. This effect often harbours more weak frequency components that may be present in the data but is not distinguishable from a stronger component. This may interfere with the spectral analysis of the signals that are present.

The FFT algorithm has N finite samples to its input and N finite samples as output, with a sampling frequency f_s on the sequence $x(n)$. The input signal has a frequency, and if this frequency is an integer multiple of f_s divided by N, the number of samples, the energy will be contained to the spectral component of that frequency in the spectrum, while remaining frequencies components will have zero energy present. However, should the remainder not be an integer, the rise of spectral leakage will appear throughout the spectrum, meaning that some frequency components will leak over and affect nearby frequencies [12][15]. The frequency components of the time domain signal will appear as main lobes in the spectral analysis. The width of a main lobe describes the spectral resolution and the ability to tell apart different components. However, as the main lobe width narrows, the energy is forced out on its side lobes, generating leakage. The effect is somewhat due to the default data windowing of the sequence. A window function frames spectral data in a specific time interval, shaping the original time domain data to appear in a certain way. There are different types of window functions, each with their own characteristic and purpose. A window covers a predetermined

interval, which inside remains unchanged whereas everything outside of the interval becomes zero or rapidly descends towards zero [13]. Whenever performing a FFT, a rectangular window will be applied to the result. A rectangular window only applies a finite time interval to the signal and is therefore the same as not using any window function at all. Windowing can help with containing the leakage within reasonable levels but are unable to eliminate it completely. The aim of the windowing is flattening of the amplitude in the beginning and end samples of the signal in the window to achieve better spectral stability [14]. All of the following window functions are defined for $n = 0 \dots N-1$. One of the most common window functions is the rectangular window, which takes N values and transforms the rest of the values in the data set to zero. However, rectangular windows often cause more spectral leakage than desired as they do not perform any adjustments to amplitude in the side lobes. Rectangular windows are favourable when it comes to achieving high spectral resolution, as they have the most narrow main lobe [25]. A rectangular window, seen in Fig. 2.2, has the value one over its entire length and is therefore defined by the equation [11]

$$w(n) = 1 \tag{2.5}$$

A different common window function type are characterized by a sinusoidal shape, such as the Hann and Hamming windows [20]. They both have a wide peak at the center of the window interval, with degeneration towards the sides. The Hann window function is described by

$$w(n) = 0.5 - 0.5cos\frac{2\pi n}{N} \tag{2.6}$$

The Hamming window, seen in Fig. 2.3 is described by

$$w(n) = 0.54 - 0.46cos\frac{2\pi n}{N} \tag{2.7}$$

Multiplying a sampled sequence of the time domain sequence with the specific window function generates a new frequency spectrum characteristic. These windows mitigates the spread of the leakage and centers it around most dominant frequency components of the original signal. An example of this is shown in Fig. 2.4 where one can see how a window helps to confine the spread of energy over the spectrum. The advantage of windows is not centered around replicating the signal identically to the input. The largest benefit lies with the confinement windows provide to the spread of the leakage over the frequency spectrum. Fig. 2.4 exhibits how the window function can confine leakage to not spread far from the main lobe.

2.4 Welch's method

Welch's method is a PSD estimator approach, which calculates the power of the desired signal at certain frequencies, allowing for spectral analysis. The method is preferred over standard methods of time to frequency conversion as it tends to reduce noise levels in the final results at the cost of frequency resolution. To

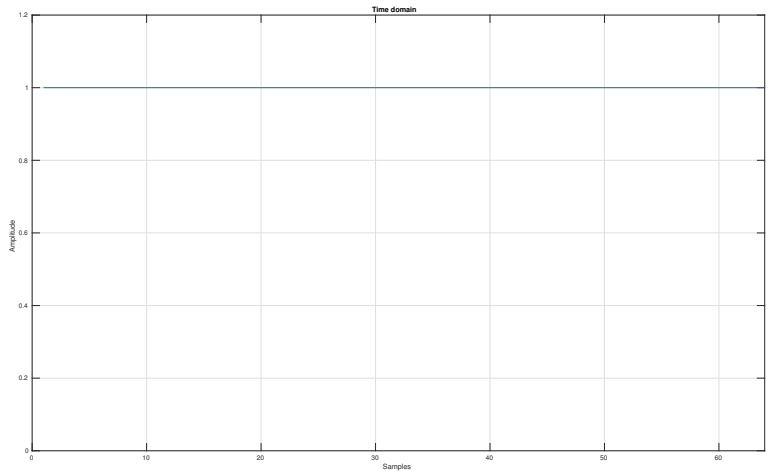


Figure 2.2: Example of the rectangular window.

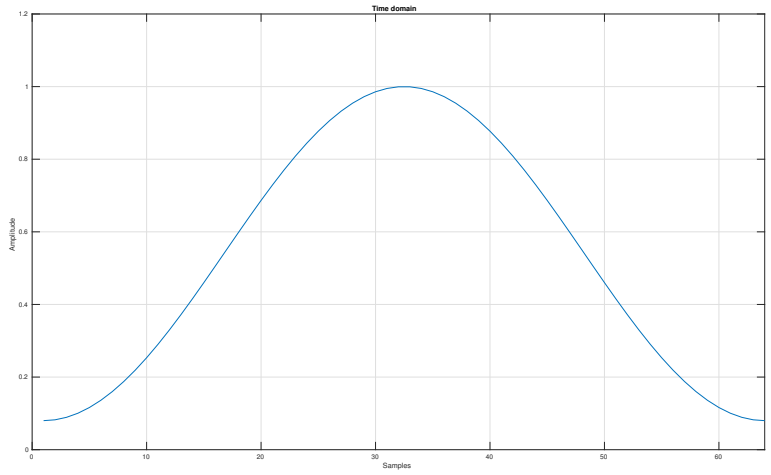


Figure 2.3: The Hamming window, which is the default window for Welch's method.

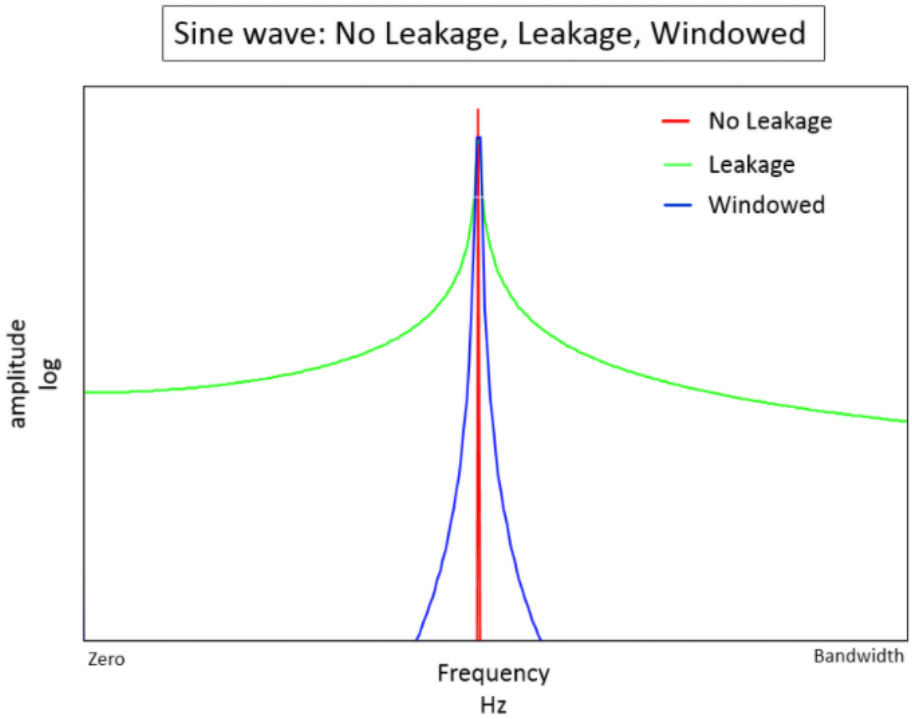


Figure 2.4: The red curve shows a periodic sine wave without any leakage in its frequency domain. The green curve represents a non-periodic sine wave with leakage and the blue curve is the same sine wave but with a window applied to it [31].

apply Welch's method, the signal in its time domain representation are split in to a number of segments, possibly overlapping each other. The size of each segment is chosen based on if higher frequency resolution or further reduced noise is desired. Larger segment sizes will reduce the total number of segments which reduces the amount of averaging between the segments, which in turn will not be as effective at eliminating noise but will yield a higher frequency resolution since the FFT is performed using more samples. Using smaller segment sizes will instead increase the total number of segments, giving the algorithm more data to average, which reduces the noise levels but with a lower frequency resolution [24]. The segments are then subjected to windowing, a function which frames data in subsets of larger data sets, for reducing spectral leakage as explained above. For each segment, the windowed FFT is computed and used to calculate the periodogram. The method then averages all of the values in the periodograms, forming the Welch's PSD. [18] The window type denotes the appearance of the FFT. First applying the window function to the input signal, and thereafter calculating its FFT. Therefore the FFT will attain a different shape depending on the specific window function. Consider the system *H(z)* with the input signal of *X(z)*. The output *Y(z)* can then be decided by [19]

$$Y(z) = X(z)H(z) \tag{2.8}$$

The specific window transfer function *W(z)* can be applied to the input signal which results in

$$\hat{X}(z) = X(z)W(z) \tag{2.9}$$

The new windowed signal can now be used as the input to the system, generating the new output

$$\hat{Y}(z) = \hat{X}(z)H(z) \tag{2.10}$$

Welch's method allows for customization so that the user may choose which window to apply to the segments. The m:th segment of the signal $x(n)$ with the chosen window function *w(n)* applied to it is

$$x_m(n) = w(n)x(n + mR) \tag{2.11}$$

where $n = 0 \dots M - 1$ and $m = 0 \dots K - 1$ and where K is the total number of segments in the signal x which also denotes how many periodograms are averaged together to form the final Welch's PSD estimate. R is the window hop size which denotes by how many samples each window will advance forward in time [21]. As each segment becomes windowed with the previous equation the periodograms for each segment can now be computed. The periodogram P for the specific m:th segment can be calculated with [16][17]

$$P_{x_m}, M(\omega_k) = \frac{1}{M}\left|FFT_{N,k}(x_m)\right|^2 = \frac{1}{M}\left|\sum_{n=0}^{N-1} x_m(n)e^{j2\pi nk/N}\right|^2 \tag{2.12}$$

where M is the amount of data samples in the original signal. Now the periodograms can be averaged together, resulting in the Welch's PSD estimate [22]

$$S_x(\omega_k) = \frac{1}{K}\sum_{m=0}^{K-1} P_{x_m}, M(\omega_k) \tag{2.13}$$

As described before, Welch's method creates frames or segments of the signal. Each segment will overlap the nearby segments, if not a rectangular window function is used. This effect will cause a redundancy in information at the places the frames overlap. By using a window that is not rectangular, one may avoid this as they are less prominent towards the edges of the segment, that is, where they overlap. A central trade off lies in Welch's method between spectral resolution and statistical stability, that is low variance over the averaged data. The size of each segment M is desired to be maximized in order to achieve a high spectral resolution. However, increasing the number of segments K will cause more averaging to occur resulting in a better stability of the analysis. Default choice of segment size is usually $M \approx K \approx \sqrt{N}$, where N is the number of data samples [23].

2.5 Equalization

An equalizer is an audio device employed when attempting to balance frequency amplitudes in a signal to a certain level. The EQ alters the gain on particular frequency bands, resulting in a new sound profile. The EQ can be used to assure some sounds appear more prominent or to be completely eliminated as well as distinguishing certain aspects of a signal, for example the bass [8].

The EQ found in the loudspeakers used in this project are of parametric type, which entails that they are susceptible to change on three different parameters: amplitude levels, center frequency and the bandwidth. The center frequencies allows for a more accurate adjustment as one can specify exactly which frequencies to change. The number of center frequencies is decided by how many bands the specific EQ contains. The parametric EQ may also decide in what extent it will affect neighbouring frequencies of the center frequency through the Q factor which is calculated using

$$Q = \frac{f_0}{B} \tag{2.14}$$

where f_0 denotes the center frequency for the band and B the bandwidth which can be calculated using

$$B = f_2 - f_1 \tag{2.15}$$

where f_1 and f_2 denotes the lower and upper crossover frequencies, respectively, meaning where the gain is attenuated by 3 dB, or where 50% of the power is lost. Based on equation 2.14, a larger Q makes the EQ affect a smaller range of nearby frequencies whilst a more narrow Q will affect more frequencies since B will be larger [10].

2.6 Different microphones

There are different types of microphones, excluding common disparities as frequency response and sensitivity. Two common types of microphones are the unidirectional and omnidirectional microphones.

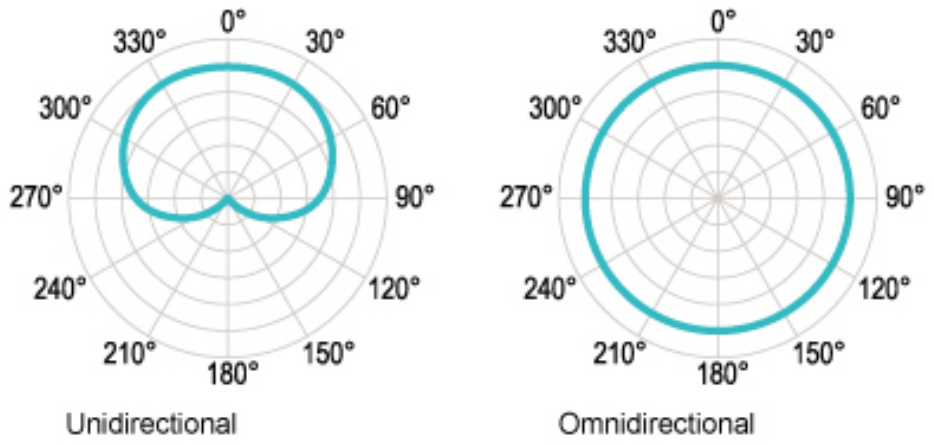


Figure 2.5: Polar pattern of a unidirectional microphone (left) and an omnidirectional microphone (right) [7].

2.6.1 Unidirectional

The directional or unidirectional microphones has the highest sensitivity and gain in its facing direction. It can be split up in to several variations such as cardioid, hypercardioid and supercardioid. They differ in the way they reject incoming sound from different directions and typically repel or attenuate all sound coming from the rear. The cardioid is the most common type and has a "heart-shaped" recording pattern, meaning it picks up all incoming sound from the front but gradually becomes less effective the more the sound source is rotated towards the sides with its lowest microphone gain at the rear. The coverage angle is approximately 130° as seen by Fig. 2.5 and for incoming sounds within this angle the frequency response generated will be fairly flat. Although directional microphones generate higher noise levels at higher frequencies, they are advantageous in that they reduce the noise levels in the direction they are facing and also blocks out far field noise better than omnidirectional microphones, by about two-thirds. [5]

2.6.2 Omnidirectional

Omnidirectional microphones will capture sound with equal sensitivity and gain from all directions, it will however become increasingly directional the higher the frequency. Due to the physical attributes of the microphone body, such as shape and size, the microphone's possibility to attain its omni qualities at higher frequencies varies. The microphone body has a tendency to block frequencies with a wavelength that measures up to the diameter of the microphone. At high frequencies, the sound coming from the rear will typically be flattened and therefore the microphone will act closer to a directional one. Thus, it is preferable with a small body diameter in order to keep the omnidirectional characteristics when reaching very high frequencies with a typical short wavelength [41]. The circular polar pat-

tern of an omnidirectional microphone is displayed in Fig 2.5. Omnidirectional microphones are more prone to record unwanted noise, reverberations and echo to a larger extent than directional microphones as they record equally from all directions. [2][3][4]

2.6.3 Proximity effect

One major difference between a directional and omnidirectional microphone is that the omnidirectional microphones does not generate the proximity effect which typically occurs when the directional microphone is within 20 cm of the sound source. The effect causes an increase in the lower frequencies, creating a stronger bass effect. This is because as the distance between the microphone and the sound source decreases, the resulting sound pressure rises. This results in the microphone exhibiting a different frequency response than expected. [6]

2.6.4 A, C & Z weightings

There are different ways of interpreting the incoming frequency spectrum since sound is relative to different devices and applications. The main weightings which affects how the microphone will display the frequency spectrum are called the A, C and Z weightings. Human ears are most sensitive to frequencies between 1000 Hz - 6 kHz [34]. In order for the frequency spectrum to represent how the sound actually is perceived, different frequency weightings are used in order for the microphone to compensate for the ear's nonlinear frequency response. The frequency responses for the different weightings are presented in Fig. 2.6. The magnitude of these different weightings are denoted dB(A), dB(C) and dB(Z).

A weighting

The A weighting is designed to represent the human hearing by attenuating frequencies below 1000 Hz and above 6 kHz since the ear does not perceive these frequencies with equal sensitivity. Measured sound with A weighting is denoted dB(A).

C weighting

Much like the A weighting, the C weighting represents the ear's frequency response but focuses at higher sound levels, for example 100 dB and above and is mainly used for measuring peak sound pressures since the ear's frequency response is flatter at higher sound levels.

Z weighting

Contrary to the other weightings, the Z weighting is the flat frequency spectrum where the recorded sound is not altered in any way, except for individual microphone responses.

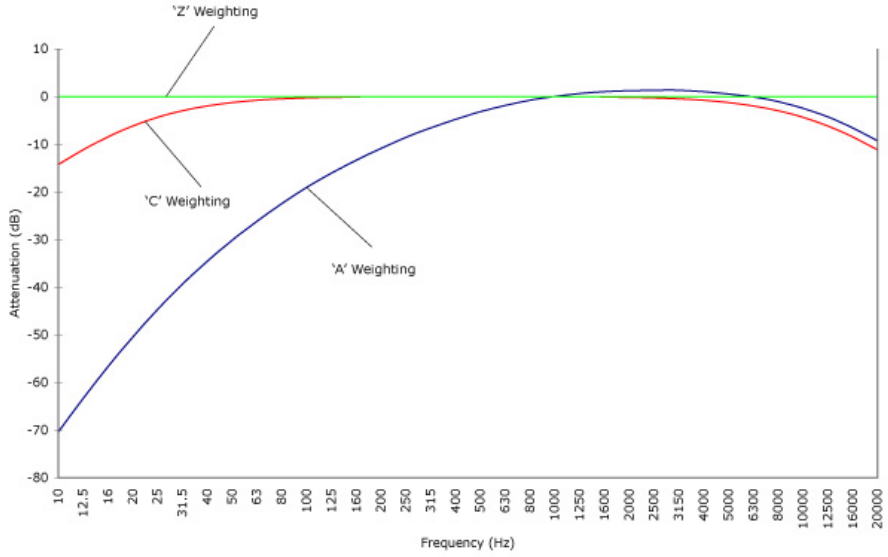


Figure 2.6: Frequency weighting for A, C and Z from 10 to 20 kHz [30].

Chapter 3

Hardware and software

3.1 Hardware

All hardware used in this thesis were supplied by Axis Communications AB. In order to be able to perform acoustic room corrections, both speakers and microphones were used. The following sections describes the hardware used in more detail and which limitations each individual component poses. At the lower level, every component in the Axis environment is using an embedded PCB running a Linux subsystem. Using the same hardware base for every device shows considerable advantages when running tests and configuration since everything can be automatized without code deviations. The devices are using PoE as their power supply, which means that the only connection necessary is the network connection used for communicating with the devices.

The system will consist of a server, a client in form of a PC program or a mobile phone application and a number of microphones and speakers. The server will perform all needed calculations for all implementations and feedback the results to the client which is the application that will display them for the user. In simple terms, the server will control all devices and execute the calibration along with other automated tests and configuration.

3.1.1 Speakers

Axis C1004-E

One of the supplied speakers is the Axis C1004-E which has a full-range speaker element with a sensitivity of 96 dBSPL. Since all Axis products have the same embedded board, the C1004-E is also an active speaker with it's own power supply and software. The C1004-E has a ADAU1761 DSP shipped with controlling software. Along with a speaker element, it also has an embedded unidirectional microphone used for calibration and testing.

3.1.2 Microphones

Brüel and Kjær Type 2250

For sound recording in an anechoic chamber a highly adaptive hand-held device from Brüel and Kjær was used. It contains a 1/2 in prepolarized microphone that is able to record sound and vibrations from 6 Hz to 20 kHz. Start and stop button allows for time restricted recordings and an internal storage easily accessible from the computer by USB connection or via network through a micro USB adapter. Real-time graphs provided in the user interface display analysis in either 1/1 or 1/3 octave bands. It is possible to choose between A, C and Z weightings for the recordings as well as a maximum and minimum dB level allowance. The microphone can be calibrated using the enclosed calibration device attached to the end of the microphone.

Axis T8353

T8353 is a small omnidirectional microphone with a frequency range of 20 to 20kHz. The microphone has been subject to testing by Axis as shown in Fig. 3.1. An omnidirectional microphone should be used for recording sound in a room since the sound waves might arrive at the microphone from different directions due to reverberations and placements. An unidirectional microphone will perform worse at recording waves arriving in a direction where the microphone is not positioned as explained in section 2.6.

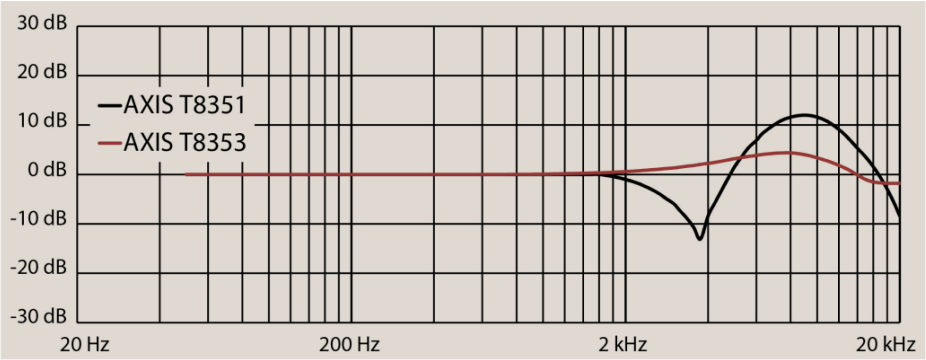


Figure 3.1: Frequency response curves for two Axis microphones. The red curve represents the model used in this project [1].

3.1.3 DSP ADAU1761

The embedded DSP software, dspd, developed by Axis supplies all functionality connected to the hardware. The dspd offers two different equalizer settings whereas only one is configurable, called static and preset. The static equalizer is configured by Axis to maximize the speaker's potential and will not be turned off by Axis request. The preset equalizer is a 1/1 octaves parametric equalizer with 9 bands

with a constant Q factor of 1. For the moment, all bands can be changed to ± 12 dB with the current firmware but may be subject to change in the future without any hardware replacements. The octave bands used for the preset equalizer is 63 Hz, 125 Hz, 250 Hz, 500 Hz, 1000 Hz, 2000 Hz, 4000 Hz, 8000 Hz and 16000 Hz. These are the center frequencies of the 9 bands in the EQ. The dspd software also provides functions to further coloring the sound, such as loudness, compressor, high-pass filter and limiter.

Loudness

Loudness compensation is used to compensate for the ear's change in sensitivity as the volume is altered, specially for the lower and higher frequencies which is based on the equal-loudness contours [39]. Playback at lower volumes will be perceived as less bass and treble heavy without taking the loudness contours into consideration. Although, in order to perceive the sound as intended during recording, the actual master gain would need to be known which is something that is not usually provided. The actual loudness compensation is based on the current gain of the output.

Compressor

Compressing is the process in which sound is attenuated in case it is exceeding a decibel threshold set by the compressor. Some common reasons for compressing audio involves smoothing out voice levels to not appear as shouting or compensating for unfavourable microphone positions. The compressing of the signal could help the human ear perceive sounds in a more subtle and comfortable way as the ear is generally sensitive to abrupt changes in energy levels. Another reason for compressing is to decrease the dynamic range of the signal. The dynamic range is the amplitude difference between the loudest and the most quiet signal. By compressing the loud signals to a certain level, the more quiet sounds can become more audible as a result, generating a more smooth sound profile. By limiting the dynamic range, the resulting RMS value may increase, which tells the average power of the voltage.

Some audio devices deals with technical limitations that does not allow them to exceed a certain peak point in the signals in order to avoid equipment damage. These devices can greatly benefit from compression. Since the compressor decreases the difference between quiet and loud sounds, the dynamic range is also reduced. [37][38]

High-pass filter

A high-pass filter is a filter which attenuates frequencies below a specified cut-off frequency. The cut-off frequency is where the attenuation reaches 3 dB. Frequencies below the cut-off frequency steeps accordingly to the filter design. The Axis speakers use this feature to prevent playing frequencies which the speakers are not able to reproduce accurately anyway, saving power. [45]

Limiter

A limiter is a type of compressor but acts less smooth in its process of attenuating the signal. It has a higher ratio of compression and may cause the signal to be distorted if the attack time is too fast, however a limiter generally prevent sudden volume peaks to appear during live sound, which could help prevent loudspeaker damage. In the case of Axis' speakers, the limiter is mainly used to make sure the audio hardware is not using too much power since PoE only supplies 15 W [35].

3.2 Software

The C1004-E loudspeaker has a specific hardware and software setup developed by Axis Communications. Since all used equipment is running the same base with only PoE connections and an embedded Linux distribution, special software needs to be written in order to be able to run the desired tests for automation and receiving results since there are no dedicated line-in or line-out connections as common passive speakers being connected to an amplifier. In order to communicate with the devices, Axis Communications is using the open source multimedia framework GStreamer which handles pairing of devices, synchronized playback among relevant features [36]. Playing audio is handled via a web interface where it is possible to pair different loudspeakers to act as a group. While playing audio from a web interface is plausible, automating the process of playing and recording audio and analyzing the available data is somewhat restrained by using the web interface. All relevant software used for testing and recording is written in C++ due to the complexity of the algorithms used and for simplicity while testing. The structure is divided into a server and client structure where the server handles all communication with the loudspeakers and microphones and the client merely acts as an interface which presents options and feedback to the user. Per Axis request, a simple, automated configuration of all devices was developed as a feature. The whole system is built with the prospect of having a passive server with the client sending requests and waiting for feedback. As per the background and thesis motivation, Axis pursued a simple solution for room correction using multiple points of interest. Using this structure is motivated by the need of automation and simplicity.

3.2.1 Server

All configuration along with playback and recording is available through shell commands by having a SSH connection to the device. The server part of the system combines all of these relevant commands and chains them together to automate testing and configuration. The server itself opens SSH connections for communicating with the devices. By chaining different commands together and keeping open connections to all relevant devices, it is possible to synchronize command execution on all devices and writing different commands to different devices simultaneously. This feature is used for recording audio in every PoI simultaneously while having the option to start audio playback in different loudspeakers.

The Axis products ships with different audio effects and alterations which needs to be turned off for accurate testing. Some of these and why they should be turned off are explained in section 3.1.3. As mentioned, all devices communicate using Ethernet which introduces problems compared to passive speaker designs where usually there's an central amplifier responsible for setting the correct settings for each speaker. Since all devices have their local settings, the server needs to configure every device manually and equally in order to have reliable results while testing.

3.2.2 Client

The client part of the system merely acts an interface towards the user, delivering feedback from the server. It enables the user to run different commands, for example executing loudspeaker localization, setting default settings for all devices, checking device availability, enabling or disabling equalization and enabling or disabling Axis own sound enhancements.

3.2.3 Android application

A mobile phone application was built for the Android OS using Android Studio and the Java programming language. The application was structured as a shell, with communicative channels such as buttons and progress bars to register the users interactive actions. Once interacted with, the application would communicate with the server that would execute the action depending on what the user requested. The application was meant to be used by a potential installer or the maintenance staff. It would allow for an easier and more flexible way of configuring and calibrating the loudspeakers rather than using a computer, as the loudspeakers would often be located in large areas requiring portability of the communication device. The Android application is meant to provide the same functionality as the PC program.

Loudspeaker localization

One functionality is to retrieve a map of the room where the mobile phone is located. This map will contain a grid view with dots to represent each speakers' coordinates. In this view, the user will have the option to press on each dot individually to see the IP of the device connected to those coordinates and the current output and microphone volume of that device. This allows the maintenance personnel to easily locate a specific loudspeaker and also get information of which loudspeakers are in close proximity of each other and thereby may be configured in the same way. The coordinates are retrieved from the server localization software that will be explained in section 4.2.2.

Calibration

When entering the calibration view from the main screen in the application the user will be met by a tab layout containing two different tabs. The initial view in the first tab contains the non-equalized input to the microphone as a bar chart graph

as well as a toolbar with buttons for all the functionality needed for equalization. As the system allows an arbitrary amount of speakers, the application has support for choosing which microphone to see the results for using a simple drop-down list. The chosen microphone will have its input bar chart revealed in the first tab and the user may change microphone at any point. The toolbar located at the bottom of the view contains several different buttons according to the list below.

- **Start calibration** - This action will open up a dialog screen where the user will have to confirm or cancel if they want to start the process, thus erasing all current settings. Pressing "Run" will start the equalization process with all connected ĺoudspeakers and microphones. Upon confirmation, the view will also change to the second tab, where the bar chart graph of the selected microphone will update itself during run time, allowing the user to see the difference with the new EQ applied to each loudspeaker. This seamless interchange between the first and second tab allows the user to determine whether the equalization was successful or not by viewing the results and the score value.
- **Stop equalization** - This button allows the user to stop the ongoing process of equalizing. The loudspeakers will not get any new EQ applied to them as a consequence.
- **Reset** - Pressing this button will reset all loudspeakers to their original EQ before start. This means that all calculated values will be lost and and the EQ will return to the values it had before any equalization process occurred.
- **Axis effects** - Toggles the Axis built-in DSP features that are explained in section 3.1.3.

List of devices

It's possible to retrieve a list of all devices connected to the network. In this list every device is clickable, opening up a new screen with further information and options, such as microphone and output volume. For individual localization purposes, it is also possible to start playing a signal from the selected speaker. One may also perform individual loudspeaker boost or attenuation by adjusting its equalizer to desired levels.

3.3 Setup

A standard setup of the hardware and software can be viewed in Fig. 3.2 where an arbitrary number of clients can connect to the server and request different commands being executed on the devices. One client could command the server to run the automated calibration with a predefined set of loudspeakers and microphones while another client could retrieve information about the devices' configuration simultaneously. The server would in that case attempt to connect to these devices using standard protocol SSH and feedback the client with any possible results. As explained above, both the mobile application and PC client communicates

with the server using the same protocol making the server unaware which client is connected.

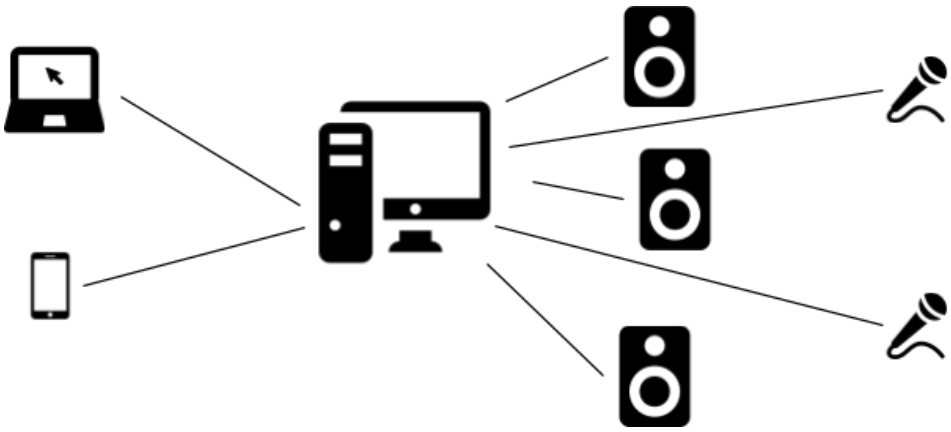
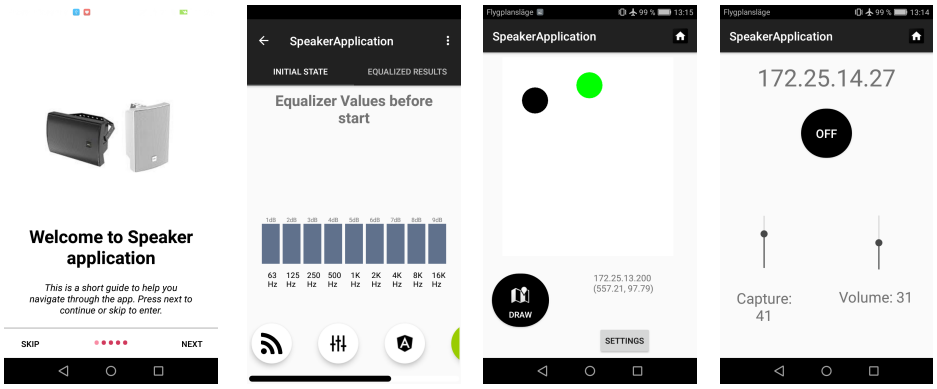


Figure 3.2: A standard setup using multiple clients and devices [40].

3.4 Android application GUI

The Android application updates the user in real time based on the server feedback. Sample outputs can be viewed in Fig. 3.3.



(a) Intro screen. **(b)** Calibration. **(c)** Localization. **(d)** Speaker settings.

Figure 3.3: Sample outputs from the Android application.

Chapter 4
Localization and distance calculation

4.1 Motivation

In a real, non-simulated scenario where a potential installer is instructed to calibrate the room's response to a chosen number of points of interest, it is vital to know that the specified loudspeakers are functional and placed as the customer suspects. In case some of the devices are misplaced, the final sound configuration could be altered in a way that decreases the overall sound experience. For example, including loudspeakers that are supposed to be placed closer to the points of interests, but instead are placed further away will reduce the quality of the final result since boosting devices that is not in range of the specified point will affect the surroundings of where the loudspeaker is actually placed. As an effect, the localization and distance calculation acts as a sanity check for the system to make sure everything is where it is supposed to be as well as providing an overview for the installer of the system. It helps the installer to ensure that the loudspeakers has been placed correctly in the room, can act as identification of a specific loudspeaker and allows for the creation of clusters of loudspeakers. Having clusters of loudspeakers grouped up together will help to configure all grouped loudspeakers with the same settings or calibrate the specified group instead of an entire area.

4.2 Theory

4.2.1 Synchronization

Usually, localization is dependent on the speakers having their system clocks synchronized in order to remove jitter and clock drift from the equation. This means that both recording sound and playing sound should be synchronized in a way such that synchronized recording should produce the same time-stamps. All Axis devices are currently only synchronized with NTP which gives sub-par clock synchronization compared to something like PTP. Since NTP does not guarantee millisecond synchronization, it is not suitable for localization since several milliseconds drifting of the system clocks will produce errors in margin of meters. Using PTP would probably offer sufficient synchronization, but requires hardware support. The current user scenario for the Axis products does not require millisecond synchronization, which is why no hardware support is present for the moment.

To produce sufficient localization approximations, the clock drift and jitter needs to be eliminated mathematically since perfect synchronization between system clocks is not possible without relevant hardware support. [32]

4.2.2 Distance calculation

Using the assumptions in section 4.2.1, the system needs to be modelled in a way that takes the synchronization problems into account in order to calculate the actual distances between the speakers. The time elapsed for speaker A to play a sound until it gets recorded and recognized in speaker B can be modelled as

$$T_{AB} = T_{p_{AB}} + \Delta_{AB} \tag{4.1}$$

Similarly, the time elapsed for speaker B to play a sound until it gets recorded and recognized in speaker A can be modelled as

$$T_{BA} = T_{p_{AB}} - \Delta_{AB} \tag{4.2}$$

where T_{AB} and T_{BA} is the total time for the sound to be played until it is registered. $T_{p_{AB}}$ is the actual time it takes for the sound to travel the distance between speaker A to speaker B. Δ_{AB} is the difference in time between system clocks, between A and B. Should A and B be perfectly synchronized, Δ_{AB} would result in being 0. Since $T_{p_{AB}}$ represents the actual time elapsed for the sound to travel between A and B, it is assumed that $T_{p_{AB}}$ and $T_{p_{BA}}$ represents the same variable. Combining equation 4.1 and 4.2 gives the equation

$$T_{AB} + T_{BA} = T_{p_{AB}} + \Delta_{AB} + T_{p_{AB}} - \Delta_{AB} \tag{4.3}$$

which results in

$$T_{AB} + T_{BA} = 2T_{p_{AB}} \tag{4.4}$$

where $T_{p_{AB}}$ can be solved by

$$T_{p_{AB}} = \frac{T_{AB} + T_{BA}}{2} \tag{4.5}$$

which is solvable since T_{AB} and T_{BA} are known variables. The actual distance the sound has travelled can be calculated using

$$s = vt \tag{4.6}$$

where s denotes the distance in meters, v the speed in $\frac{m}{s}$ and t the time elapsed in seconds. Assuming a temperature of 20° C, the speed of sound is 343 $\frac{m}{s}$, creating the equation [33]

$$S_{AB} = 343T_{p_{AB}} \tag{4.7}$$

where S_{AB} denotes the distance in meters between point A and point B.

4.2.3 Localization in 2D

Section 4.2.2 describes the theory behind how to obtain the distances between different speakers. This section explains how this information will be used for calculating a final position for all speakers, either in the form of (x, y) or (x, y, z) depending on if it is possible to assume that all speakers will be placed on the same vertical level. Performing the equations described in section 4.2.2 for all combinations of speakers specified gives all distances between all speakers.

Speaker A calculates the distance to speaker B using the equations in section 4.2.2. Let's call this distance D_{AB}. Since no information about which direction the sound was received is provided, A does not know anything about where B is actually placed, hence B could be placed anywhere on the circle C_{AB}. A's coordinates are put to $(0, 0)$ to have a known starting point.

$$(x_B - x_A)^2 + (y_B - y_A)^2 = {D_{AB}}^2 \tag{4.8}$$

The coordinates of B can be calculated with

$$x_B = x_A + D_{AB}cos(\alpha) \tag{4.9}$$
$$y_B = y_A + D_{AB}sin(\alpha) \tag{4.10}$$

where $\alpha \in [0, 2\pi)$. This is called constraining the placement of speaker B relative to D_{AB}. Adding an additional speaker C to the equation system and setting the constraints specified in equation 4.8 both to speaker A and speaker B gives the equation system

$$x_B = x_A + D_{AB}cos(\alpha_{AB}) \tag{4.11}$$
$$y_B = y_A + D_{AB}sin(\alpha_{AB}) \tag{4.12}$$
$$x_C = x_A + D_{AC}cos(\alpha_{AC}) \tag{4.13}$$
$$y_C = y_A + D_{AC}sin(\alpha_{AC}) \tag{4.14}$$
$$x_B = x_C + D_{BC}cos(\alpha_{BC}) \tag{4.15}$$
$$y_B = y_C + D_{BC}sin(\alpha_{BC}) \tag{4.16}$$
$$x_C = x_B + D_{CB}cos(\alpha_{CB}) \tag{4.17}$$
$$y_C = y_B + D_{CB}sin(\alpha_{CB}) \tag{4.18}$$

where $\alpha_{XY} \in [0, 2\pi)$. Note that $XY = YX$. Solving the coordinates for speaker B in the example equation system above gives the constraints

$$x_B = x_A + D_{AB}cos(\alpha_{AB}) \tag{4.19}$$
$$x_B = x_C + D_{BC}cos(\alpha_{BC}) \tag{4.20}$$
$$y_B = y_A + D_{AB}sin(\alpha_{AB}) \tag{4.21}$$
$$y_B = y_C + D_{BC}sin(\alpha_{BC}) \tag{4.22}$$

which yields an infinite amount of solutions since $\alpha_{XY} \in [0, 2\pi)$.

An illustration of the 2D localization can be found in Fig. 4.1. In this case

M is the current loudspeaker in which the coordinates are already known and d_1, d_2 is the distances from M to loudspeaker A and B respectively. Since there are three loudspeakers in total in Fig. 4.1, there is also a third distance, d_3, which is the distance between A and B. In this example, the circles represents the different constraints the loudspeakers have towards each other.

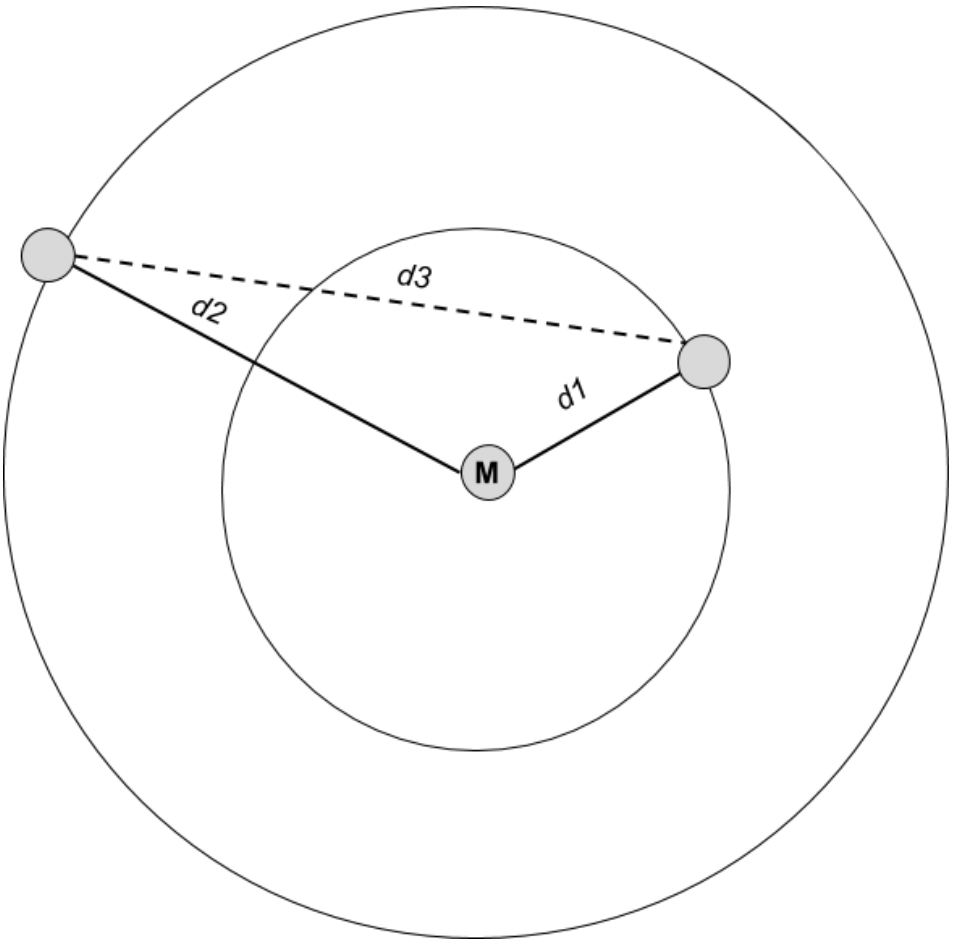


Figure 4.1: Example with a master speaker *M* and two loudspeakers at distances *d1* and *d2* away from it.

4.2.4 Localization in 3D

For locating the speakers in three dimensions, equation 4.8 can be extended to the sphere equation S_{AB}

$$(x_B - x_A)^2 + (y_B - y_A)^2 + (z_B - z_A)^2 = {D_{AB}}^2 \tag{4.23}$$

which extends the circle C_{AB} to include the third dimension as well.

4.2.5 Limitations

Localization in either two or three dimensions has one primary limitation which is the infinite amount of solutions. This can be solved with straightforward brute-forcing, by trying different values until a solution which satisfies all constraints is found. Another limitation is the fact that the rotation in neither X, Y or Z is solved. In order to overcome the rotation limitation, two speaker locations needs to be predefined for two dimensions and three locations for three dimensions in order to constraint the other speakers locations into the right rotation in space. Calculating localization based on only the distances between speakers will yield an infinite amount of solutions. One practical aspect to fix this is binding one speaker to the coordinates $(0, 0)$ for localizing in two dimensions or $(0, 0, 0)$ for localizing in three dimensions which will be used as a reference for the other speakers when calculating the other positions.

4.3 Method

4.3.1 Distance calculation

In order to realize the theory explained in the last section and to test its validity, six C1004-E loudspeakers were placed in a small room with the dimensions 3×3.5 m as shown in Fig. 4.2. All functionality was implemented in the C++ server and client explained above in an attempt to automate the testing procedure as much as possible as well as removing the human factor by allowing the tests to be started from a remote location. Measurements of the actual distances was taken prior to tests for generating an error matrix for verification.

Initially, the tests began with each loudspeaker consecutively playing a signal containing the frequency 4000 Hz while simultaneously recording everything. Since all loudspeakers will have information about when the current loudspeaker started playing as well as when the rest of the speakers started playing, the distances between each speaker could be calculated pair-wise using the formulas in the previous chapter. All recorded data was analyzed in the server using Goertzel filtering of the chosen frequency and by applying the above theoretic viewpoint.

4.3.2 Localization

In addition to distance measurements, an additional function was implemented using the theory behind 2D and 3D localization described above. The localization function takes the calculated distance data as input and outputs a plot with the locations of each loudspeaker used in the setup.

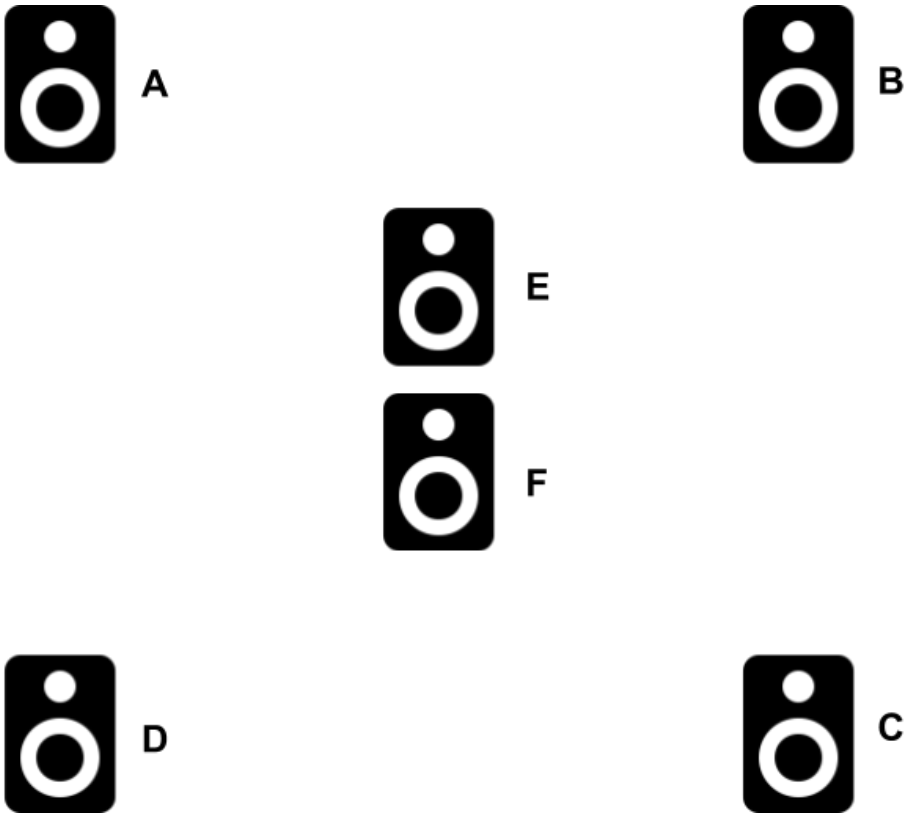


Figure 4.2: Test setup for the distance calculation and localization tests using six loudspeakers.

4.4 Results

4.4.1 Distance calculation

Table 4.1 shows the actual distances between every loudspeaker measured by hand, with an allowed margin of error of ± 5 cm due to measurement variations. Table 4.2 shows the calculated distances using the described function. Finally, Table 4.3 shows the difference between Table 4.1 and Table 4.2 in meters. The test was performed according to setup in Fig. 4.2. Table 4.3 indicates that the localization program performs adequately in comparison with the actual measured distances.

Speaker	A	B	C	D	E	F
A	0.00	2.02	2.96	2.56	1.98	2.46
B	-	0.00	2.03	3.2	2.10	2.54
C	-	-	0.00	2.15	2.19	1.90
D	-	-	-	0.00	2.45	2.09
E	-	-	-	-	0.00	0.71
F	-	-	-	-	-	0.00

Table 4.1: Actual distances between each loudspeaker in meters according to the setup in Fig. 4.2.

Speaker	A	B	C	D	E	F
A	0.00	1.97	2.96	2.50	2.10	2.48
B	-	0.00	1.94	3.26	2.22	2.70
C	-	-	0.00	2.06	2.14	1.88
D	-	-	-	0.00	2.50	1.97
E	-	-	-	-	0.00	0.72
F	-	-	-	-	-	0.00

Table 4.2: Calculated distances between each loudspeaker in meters according to setup in Fig. 4.2.

4.4.2 Localization

The localization algorithm is dependent on the distance calculations being accurate, because of this an error margin was introduced to reflect the actual validity of the result. As explained in section 4.2.5, it's necessary to have a reference point to actually compute an answer. Since the results are based on two dimensions, any height difference between the loudspeakers are ignored and instead introduces a higher error margin. Fig. 4.3 shows the actual output from the localization algorithm using the placements described in Fig. 4.2. As mentioned in section 4.2.5, the rotation is not fixed since only the distances were provided to the algorithm.

Speaker	A	B	C	D	E	F
A	0.00	0.06	0.00	0.06	0.11	0.02
B	-	0.00	0.08	0.06	0.12	0.16
C	-	-	0.00	0.09	0.05	0.02
D	-	-	-	0.00	0.05	0.12
E	-	-	-	-	0.00	0.01
F	-	-	-	-	-	0.00

Table 4.3: Difference in meters between actual distance and calculated distance using data from Table 4.1 and Table 4.2, respectively.

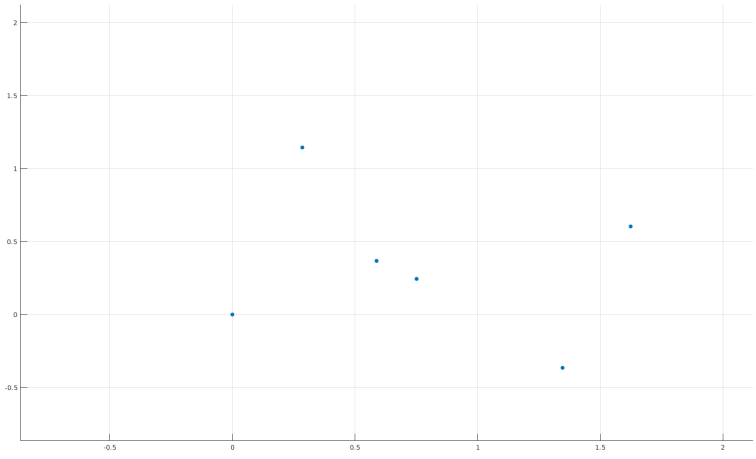


Figure 4.3: Localization results using the six loudspeakers described in Fig. 4.2.

Chapter 5
Single channel calibration

5.1 Theory

5.1.1 White noise

White noise characterizes as noise that emits the same amount of energy in all frequencies in the defined spectrum. In theory, such a distribution is physically unfeasible as it would mean there would be infinite energy since the frequency resolution would be infinite. Therefore, the signal is considered white if it generates a flat power spectrum within a specific bandwidth that covers a certain range of frequencies. Fig. 5.1 shows white noise in both frequency domain (top) as well as time domain (bottom) [44]. Other types of noise include coloured noise such as pink and brown noise. These noises have the distribution

$$P = 1/f \tag{5.1}$$

$$B = 1/f^2 \tag{5.2}$$

where P denotes the distribution for pink noise and B denotes the distribution for brown noise. This means that there is constant energy in every octave whereas Gaussian distributed white noise distributes the same amount of energy in every frequency.

5.1.2 Calibration

By playing noise from the loudspeakers, the frequency response can be measured using Welch's method described in section 2.4. The frequency response will be compared and adapted to the desired frequency response of the system. The desired response will differ depending on how the specific user wants the sound to be perceived in certain points of interest. For simplicity, a flat frequency response is shown as an example in Fig. 5.2, green curve. Each PoI, or microphone, will potentially contain a different frequency response profile. In Fig. 5.3, the red curve represents the frequency response of the current recording in the specific PoI. The blue curve represents the correction curve needed to obtain the desired frequency response, which in this example is a flat spectrum. For example, The Axis C1004-E loudspeaker implements a 9-band parametric EQ, which will be the final result of the correction curve. Since it's desired for the method to be as general as possible,

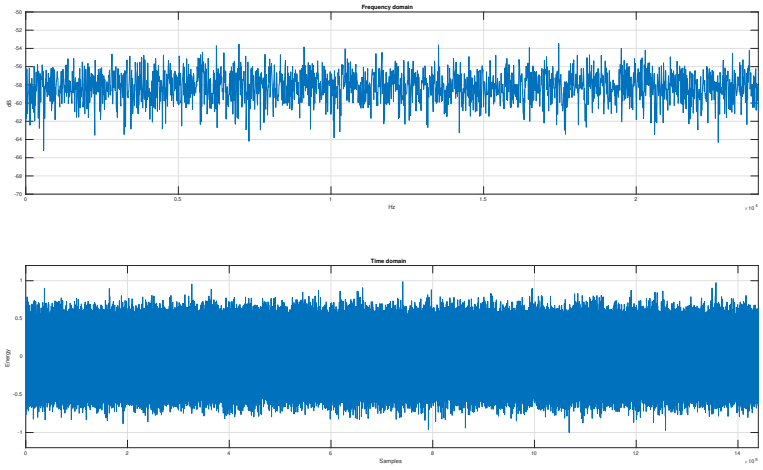


Figure 5.1: White noise represented in both the frequency domain and the time domain.

all calculations are done in high resolution before truncating the results to fit the specified EQ. In total, there will be three steps to be performed before truncating the results into the specified EQ. These steps will henceforth be called profiles. This method is described by

$$P_T = P_1 + P_2 + P_3 + C \tag{5.3}$$

where P_T represents the final non-truncated result. P_1, P_2 and P_3 are different profiles that will be applied in order to produce the final result. C represents the adjusted gain in dB. Since it should be possible to adjust both the frequency spectrum and the sound level, adding C to the final profile adjusts the sound level to the reference point specified by the PoI. The reference sound level is relative to the microphone calibration and is specified in dBFS. As the customer might request different sound profiles for the various PoIs, the loudspeakers will receive a different C depending on which PoI they affect. That is, should the microphone record a certain sound level that differs from the desired sound level, the loudspeakers will receive a specific C to compensate for the difference between the actual recorded gain and the desired recorded gain. C will not affect the frequency spectrum since it is only a change in gain. The desired level for each PoI are provided as input to the system.

Profile P_1

P_1 denotes the difference between what the microphone recorded, P_i, and the desired profile, P_f, which in our case is the flat spectrum. Fig. 5.2 shows an example of P_i and P_f. The blue curve in Fig. 5.3 displays the inverse of P_i, which combined with P_i will result in P_f. In simple terms, P_1 is the correction curve

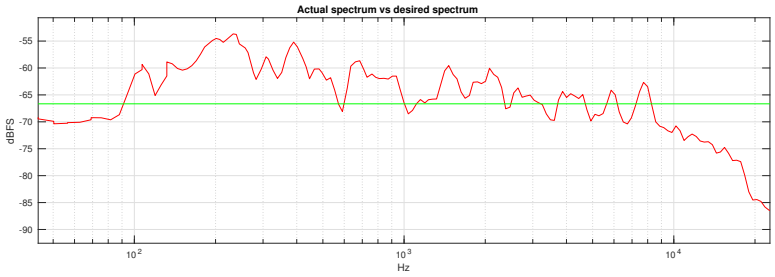


Figure 5.2: P_i and P_f example comparison. The green curve represents the desired frequency response and the red curve represents the measured frequency response in the PoI.

needed to reach P_f. Both P_i and P_f are measured in dBFS, where P_1 is only the difference between them specified in relative dB as the right Y-axis shows.

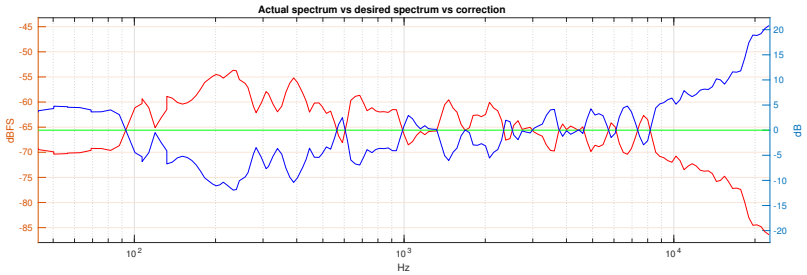


Figure 5.3: Example of P_1 where the red curve represents the measured frequency response P_i at the PoI and the blue curve represents the correction profile P_1 needed to reach the green target curve P_f.

Profile P_2

P_2 represents the physical hardware limitations of the equipment used, for example the frequency range of both the loudspeaker and the microphone. P_2 is an optional step which controls how roughly the method will apply the correction adjustments on parts of the frequency spectrum where the equipment is limited. For example, using the loudspeaker Axis C1004-E and the microphone Axis T8353 will produce a hardware profile that can be seen in Fig. 5.4. The graphical representation will provide a decent indication of where the cut-off frequencies are specified using the data-sheet for the C1004-E and T8353. As shown in Fig. 5.4, the hardware is capable of performing around roughly 100 Hz to 17 kHz before attenuating the gain due to hardware limitations.

Combining the hardware limitations with P_1 shown in Fig. 5.4 yields the profile P_2 displayed in Fig. 5.5, where the magenta curve represents the applied hardware

profiles and the blue curve represents P_1. Examining the magenta curve, it can be observed that applying the hardware profile in this case attenuated the correction magnitude for the specified hardware limited frequencies mentioned above. The correction magnitudes between the lower and upper hardware limitations are not adjusted since the hardware performs adequately within this range. Since boosting the frequencies outside of the hardware limited frequency range is unlikely to make any difference to the final sound profile, P_2 prevents unnecessary boosting which results in lower power consumption since the loudspeakers does not have to output as much energy.

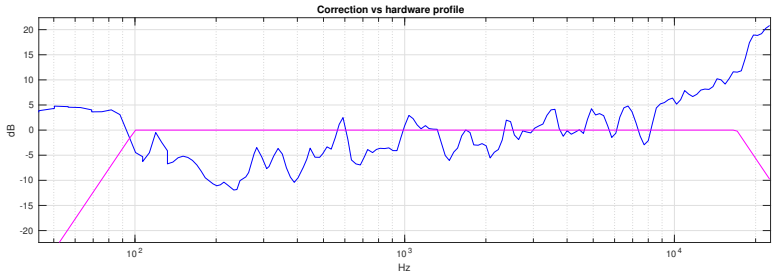


Figure 5.4: P_2 is the magenta curve and the blue curve represents the correction curve to reach the flat response P_f.

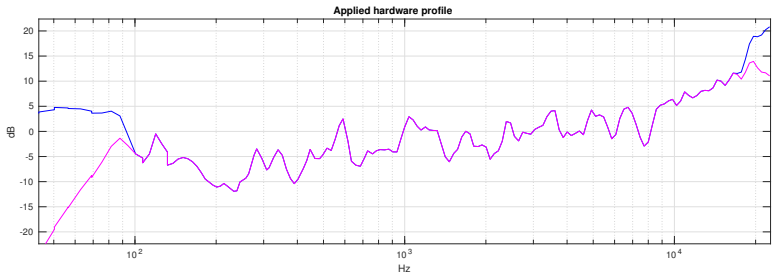


Figure 5.5: Example of adding the hardware profile P_2 to the correction curve P_1. The magenta curve represents how the resultant curve will look like and the blue curve represents P_1.

Profile P_3

P_3 is an optional step to further colour the sound based on a specific customer profile, if desired. It is plausible to think that some customers would want certain frequencies boosted, for example the frequencies responsible for the bass and treble. If the customer has no request for further colouring, P_3 will result to zero and combining P_2 and P_3 will still produce P_2. Applying a P_3 profile, will result in a new frequency spectrum appearance which is achieved through a change in EQ settings.

Profile P_T

P_T represents all of the profiles combined and represents the final adjustment. In order to apply P_T to the specified hardware, P_T is truncated in different ways. The C1004-E's 9-band EQ only allows for adjustments in 9 different center frequencies. In this case, P_T is truncated to the 9 center frequencies in the C1004-E's EQ and adjusted to fit the final compensation curve. Fig. 5.6 displays an example of truncating P_T into a specific profile, which in this case represents the C1004-E's 9-band EQ. Each blue dot along the curve represents the boosting or attenuation for each particular frequency band in the EQ.

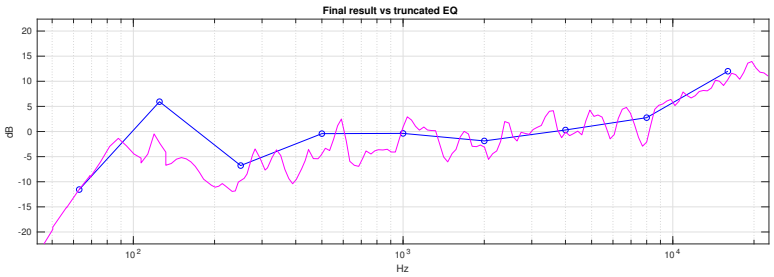


Figure 5.6: The final compensation curve. The magenta curve represents P_T and the blue curve represents the final EQ settings truncated from P_T.

5.1.3 Correction factor

With the intention of adjusting the EQ as precisely as possible, there was a need to measure the correlation between EQ settings to the DSP and the output from the loudspeaker to detect any anomalies. For example, boosting a specific center frequency with x dB should produce a boost of x dB in the recorded signal at the reference microphone of that frequency. Should the previous assumption prove to be false, a correction factor would need to be applied to the EQ settings. This can be modelled as

$$C_f = \frac{y_f}{x_f} \tag{5.4}$$

where x_f and y_f are the adjusted gain for the frequency f on the DSP and the corresponding recording gain at the microphone, respectively, and where C_f represents the quota between the actual gain and the recorded gain at frequency f. It is assumed that $C_f \leq 1$ since boosting a certain amount of energy should not produce any additional energy beyond that. Should C_f deviate from 1, it is considered an anomaly. Since the hardware might perform differently emitting different frequencies, it is plausible to introduce a correction factor for every EQ band, denoted as $C_{f,b}$ where the quota between the actual gain and the desired recorded gain is measured in the center frequency for band b. If any band is considered an anomaly, measures have to be taken into account when adjusting those specific bands to eliminate the difference between desired change and actual change. In

other words, if a correction factor is needed, the actual EQ setting $x_{f_{new}}$ can be modelled as

$$x_{f_{new}} = \frac{1}{C_{f,b}} x_f \tag{5.5}$$

where $x_{f_{new}}$ denotes the adjusted EQ setting and x_f the desired gain.

5.2 Method

One test were conducted in an anechoic chamber in the E-building at LTH. The anechoic chamber is displayed in Fig. 5.7. An anechoic chamber is used as a reference for further testing since any reverberations will not be present, making the sound propagate the same way for every test. Every test presented in this section, excluding the distance attenuation test, was conducted using one loudspeaker, Axis C1004-E, and one microphone, Axis T8353.

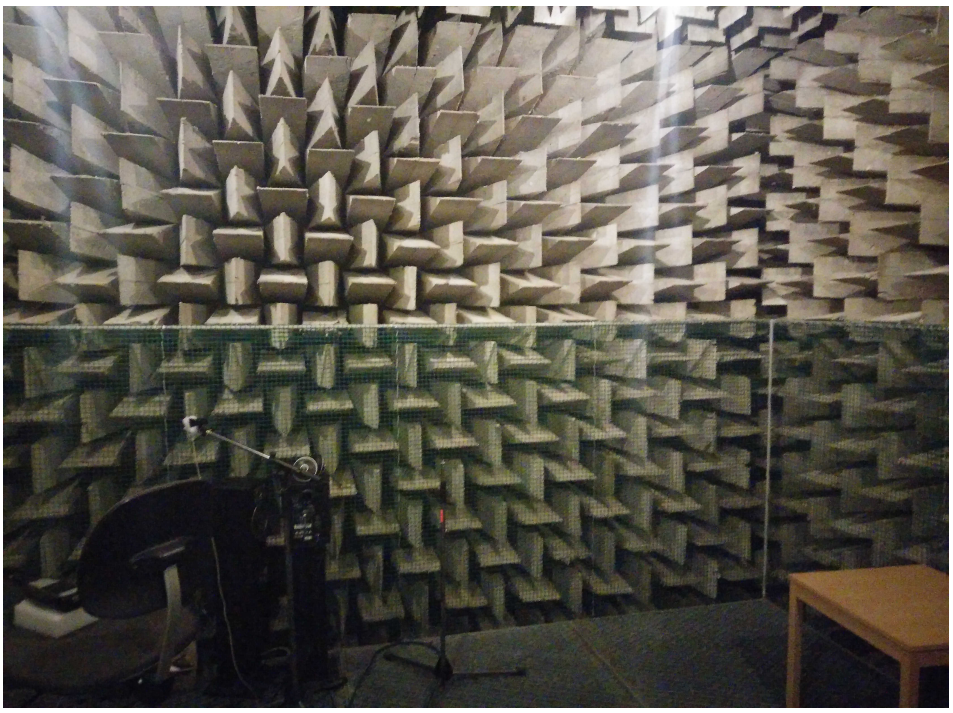


Figure 5.7: Anechoic chamber in E-building at LTH.

5.2.1 Distance attenuation

To measure how the sound dissipated with distance, a distance attenuation test was formed for validating the multi channel calibration model presented in section 6.1. The distance attenuation test was performed by placing the loudspeaker in one of the corners and measuring the sound pressure of the diagonal towards the

opposite corner. White noise was played as the measurement signal. Instead of using the Axis T8353 microphone, this test was conducted using the Brüel & Kjær reference microphone mentioned in section 3.1.2. A reading of the sound level was recorded at 1 m, 2 m, 3 m, 4 m and 5 m.

5.2.2 Calibration

For the single channel calibration, two different types of tests were conducted. The first test was conducted by playing white noise for 30 seconds to ensure that enough energy has been recorded for accurately measuring the frequency spectrum using Welch's metod described in section 2.4. The second test uses a signal containing all the center frequencies of the specified EQ, in this case the DSP in C1004-E. Since the signal only contains the 9 center frequencies, there is no need to play the signal for 30 seconds as in the first test due to the fact that each frequency present will have enough energy for accurately calculating the magnitudes using Goertzel filtering. Instead, the signal is played for 3 seconds for a quicker calibration. All tests have summarized in Table 5.1. The 9 frequency signal will be denoted tonal signal in the rest of this project.

Signal	Length	Microphones	Loudspeakers
Tonal signal	3 seconds	1	1
White noise	30 seconds	1	1

Table 5.1: Single channel calibration tests.

For every test, a 30 second sequence of white noise was recorded before and after the test to produce an overview of the sound image before and after the calibration. Choosing a 30 seconds white noise signal was a trade off between spectrum mean averaging to ensure enough energy was received versus output gain. Both changing the signal duration as well as the output gain would affect the correct spectrum mean averaging. A duration of 30 seconds was a good trade off since the gain used did not damage the loudspeaker and the microphone received enough energy to estimate the spectrum accurately using Welch's method. The following list sums up the test procedure:

1. White noise - 30 seconds.
2. Specific signal according to Table 5.1.
3. White noise for verification - 30 seconds.

After the specific signal was recorded, the recording was analyzed and new EQ profiles were suggested using the profiling theory described in section 5.1.2. Since the theory in section 5.1.2 assumes that the full frequency spectrum according to the hardware limitations is measured, the tonal signal was analyzed using Goertzel filtering in the center frequencies and the difference to P_f was set as the final EQ settings to the loudspeaker's DSP. The frequency response of the first white noise sequence was compared to the white noise verification, which represents the calibration result. Results of these tests can be seen in section 5.3.

5.2.3 Correction factor

To test for any anomalies in the loudspeaker's output, the correction factor described in section 5.1.3 was tested using the same tonal signal used in section 5.2.2 since it contains all center frequencies for the C1004-E's DSP EQ. The test was modelled as an iterative process where every iteration would apply a boost of 3 dB for each EQ band, for a total of 4 iterations excluding the reference iteration. For every iteration, the magnitude of each EQ center frequency was measured and compared against the previous value. Hence, the desired value was 3 dB in this test as described in section 5.1.3. This iterative process was repeated 100 times to ensure that enough data was collected for averaging. The testing procedure is described in more detail in Table 5.2. The data from each iterative process was then

Boost	Recorded	CF
0 dB	S_0	reference
3 dB	S_3	$\frac{S_3}{S_0+3dB}$
6 dB	S_6	$\frac{S_6}{S_3+3dB}$
9 dB	S_9	$\frac{S_9}{S_6+3dB}$
12 dB	S_{12}	$\frac{S_{12}}{S_9+3dB}$

Table 5.2: Correction factor tests describing one full iterative process, where S_x denotes the sound level recorded for x boosting.

averaged with the other repetitions to calculate a mean value for the correction factors. The standard deviation of all samples was calculated as well to validate the results.

5.3 Measurements and results

5.3.1 Distance attenuation

Performing the distance attenuation test yielded a result where the sound level attenuated approximately 3 dB per meter, or in terms of linear power, 50% lost for every meter. Detailed results are presented in Table 5.3.

Distance	Sound level
1 m	84 dB
2 m	81 dB
3 m	76 dB
4 m	74 dB
5 m	71 dB

Table 5.3: Results of the distance attenuation test.

5.3.2 Calibration

Anechoic chamber

In the anechoic chamber the setup was one loudspeaker and one PoI. The frequency response figures demonstrates pre- and post-calibration curves in the same figure with the red curve exposing the frequency response before any adjustments and the green curve representing the frequency spectrum after applying new EQ settings. Tests in the anechoic chamber was done in three ways: calibrating using white noise and the tonal signal and adjusting to a certain sound level in dBFS. The EQ settings the loudspeaker received after all profiles were applied can be seen in Fig. 5.10 and Fig. 5.11.

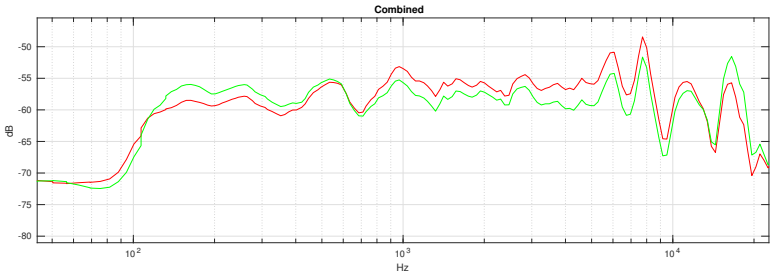


Figure 5.8: Frequency response in PoI using white noise signal as calibration method.

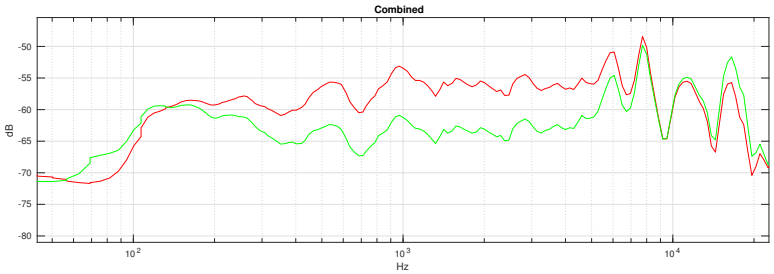


Figure 5.9: Frequency response in PoI using tonal calibration.

In order to verify that the results were efficient, the SD was used as measurement in both the pre- and post-calibration. The SD σ was calculated using the mean value from each octave band for the specific PoI, specified in dB. As seen in table 5.4, three different gain values are presented. The first G_{before} represents what the current gain was before any changes were made to it. G_{target} shows to what level the PoI was calibrated for and G_{after} is the resulting gain value that the microphone received post-calibration in dBFS.

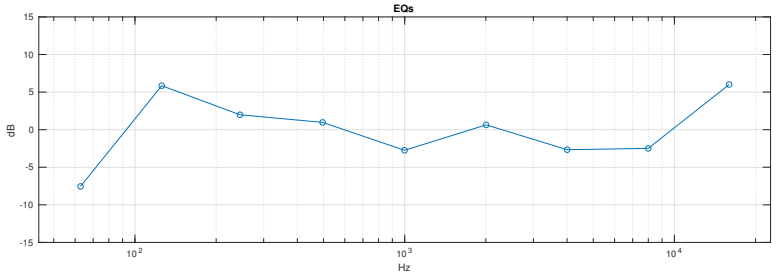


Figure 5.10: EQ values applied to the loudspeaker using noise calibration.

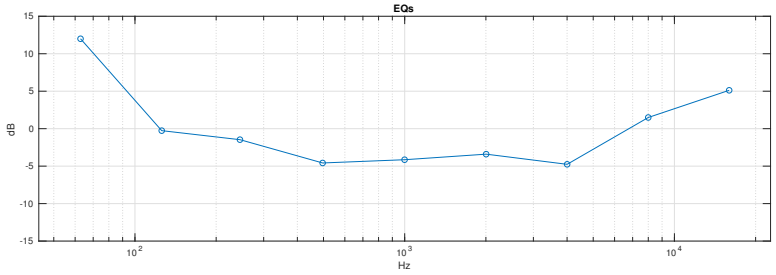


Figure 5.11: EQ values applied to the loudspeaker using tonal calibration.

Test	PoI	σ_{before}	σ_{after}	G_{before}	G_{target}	G_{after}
White noise	A	2.46	0.20	-19.62	-30	-30.22
Tonal signal	A	2.46	2.84	-19.62	-30	-34.59

Table 5.4: Standard deviation before and after calibration as well as current gain, anticipated gain and resulting gain.

Reverberation chamber

A single channel setup was also tested in a normal room with reverberations. The test with one loudspeaker and one PoI in this room was performed using only the white noise calibration. Table 5.5 shows results from the test.

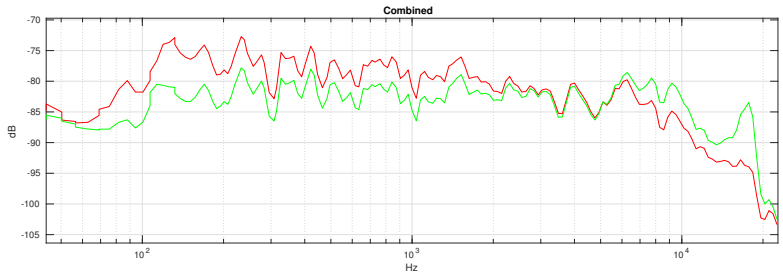


Figure 5.12: Frequency response in PoI using white noise as calibration method.

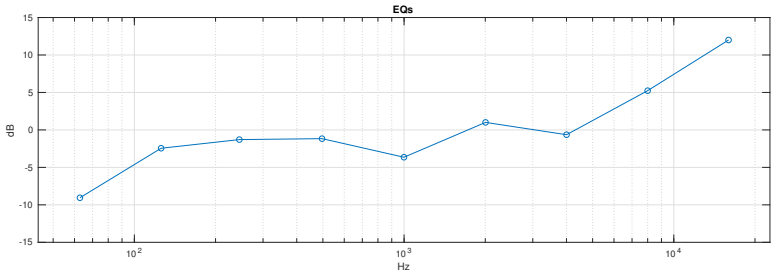


Figure 5.13: EQ values applied to the loudspeaker using white noise as calibration method.

PoI	σ_{before}	σ_{after}	G_{before}	G_{target}	G_{after}
A	5.93	2.97	-46.67	-45	-44.87

Table 5.5: Standard deviation before and after calibration as well as current gain, anticipated gain and resulting gain.

5.3.3 Correction factor

Since the correction factor test was repeated 100 times to form an accurate average, both the mean and SD was calculated. For all the 9 bands the C1004-E's DSP offers, the CF for each band implied a convergence towards 1, except for the lowest 63 Hz band. Detailed results from the CF test is presented in Table 5.6. Frequencies below approximately 70 Hz are not reliable in terms of recording due to equipment limitations and poor shielding, which the SD of band 63 Hz shows.

Band	CF	SD
63 Hz	0.87	0.74
125 Hz	0.95	0.07
250 Hz	0.98	0.11
500 Hz	0.98	0.08
1000 Hz	0.97	0.07
2000 Hz	0.98	0.09
4000 Hz	0.87	0.17
8000 Hz	0.98	0.02
16000 Hz	0.98	0.07

Table 5.6: Results of the CF calculation tests.

Chapter 6
Multi channel calibration

6.1 Theory

6.1.1 Interference

Interfering waves is the phenomenon where two waves with correlated frequencies or sound sources are combined forming a resultant wave of either greater, lower or the same amplitude depending on their phases relative to the other. Wave interference is observed with all types of waves, including sound waves. The amplitude of the resultant wave is entirely dependent on the relative phase between the waves. When the phases are in synchronization, the resultant amplitude will result in the two waves combined, also called constructive interference. When the phase is shifted by π, the wave is cancelled out, called destructive interference. In other cases, combining waves results in vector addition. Combining two sounds with the same phase will create a resultant wave with the amplitude of the two waves combined. [43]

6.1.2 Calibration

Compared to the single channel calibration, calibrating the system using an arbitrary amount of loudspeakers and points of interest introduces additional problems that needs to be taken into account. Based on the theory about wave interference in section 6.1.1 and how it will affect certain frequencies differently in certain points, the calibration should not be performed by playing the calibration signal simultaneously in all loudspeakers. Doing so introduces non-constant wave interference due to the fact that playing noise will make the loudspeakers cancel each other out in certain frequencies, and cause constructive interference in other frequencies. The non-constant error originates from the lack of synchronization between the different loudspeakers due to the network-distributed buffering of sound samples. Different synchronization in every test will lead to different interference in different frequencies which won't represent the actual performance of the calibrated room. In order to overcome this problem, every loudspeaker plays the calibration signal consecutively to eliminate the effect of wave interference [43]. Another problem with the multi channel calibration is that different points of interest will, probably, request different equalization profiles for every loudspeaker since the transfer

function for the room will be different depending on where the PoI is located. This will be taken into account and a possible solution is laid out in the next section.

6.1.3 Equalization distribution

When utilizing multiple PoIs in the tests, each microphone will receive a certain amount of energy from each loudspeaker in the room. To acquire a flat frequency response in the PoIs, one must therefore adjust the loudspeaker equalizer according to how much energy every PoI received from each loudspeaker. Loudspeakers in proximity to a specific PoI should be affected more by the close point than PoIs further away since the final equalizer settings for that specific loudspeaker will not propagate as well to those points further away as shown by the distance attenuation test in section 5.3.1. The distance attenuation test shows approximately 3 dB/m attenuation, which translates to 50% of the power lost for each meter the PoI is moved from the speaker [42].

Placing two speakers 1 m and 2 m respectively from a PoI would therefore show that the PoI would pick up half the energy from the loudspeaker furthest away compared to the close speaker. The total amount of energy received would therefore translate to 100% from the close loudspeaker and 50% from the loudspeaker further away. If the total energy received is normalized to [0, 1], this case would be 0.66 from the close loudspeaker and 0.33 from the speaker further away. This principle is repeated for every available equalization band since different loudspeakers can be positioned in different ways, boosting certain frequencies even though they are positioned further away than other loudspeakers. This repetitive procedure is modelled as

$$S_{i,b} = \sum_{j=1}^{n} M_{j,b} \frac{L_{j,b}}{L_{total,b}} \tag{6.1}$$

where S denotes the final EQ setting for band b, i the current speaker, M_j the current microphone, b the current equalization band, n the total amount of microphones and L_j the received energy from loudspeaker i to microphone M_j and L_{total} the total amount of energy received in every microphone from loudspeaker i.

6.1.4 Gain distribution

As specified in section 6.1.3, the settings requested by the PoIs should be weighted against each other since all changes will affect all points. Different PoIs might request different sound levels as specified by the problem formulation. Section 6.1.3 models a repetitive procedure to weight the different EQ settings specified by the PoIs. Adjusting this model to apply for distributing gain can be modelled as

$$G_i = \sum_{j=1}^{n} M_j \frac{L_j}{L_{total}} \tag{6.2}$$

where G denotes the change in gain, i the current loudspeaker, M_j the current microphone, n the total amount of microphones, L_j the received energy from

loudspeaker i to microphone M_j and L_{total} the total amount of energy received in every microphone from loudspeaker i.

6.2 Method

In the multi channel calibration, different setups of loudspeakers and microphones was used. As with the single channel calibration, two different signals was used as different calibration methods where white noise was played before and after the calibration for verification, as described in section 5.2.2. Many tests were conducted, where different parameters was adjusted. The parameters included amount of loudspeakers and PoI, calibration signal, test environment and sound level in different PoI. Table 6.1 describes the setups and the results can be found in section 6.3. The tests selected for the result section was representative for the method's performance and therefore showing frequency responses from all tests would be redundant and has been omitted. The specific signal in Table 6.1 was

Room	Signal	Loudspeakers	Microphones
10.5 m × 7.5 m	White noise	4	4
	Tonal signal	4	4

Table 6.1: Test setup for the multi channel test.

played consecutively from each loudspeaker in order to avoid interference. Lastly, the white noise signal was played again as verification as mentioned before.

6.3 Measurements and results

6.3.1 Calibration

The multi channel calibration had a setup according to Fig. 6.1. Four loudspeakers and four microphones were used. Two tests were conducted, one using the tonal signal for calibration and one using white noise for calibration. In the setup, PoI H and G was placed in the center of tables and chairs, which is heavily utilized during lunch and break hours. PoI E was placed in front of the elevators and the staircase, where many people pass by during a normal day. The last microphone, PoI F, was placed at the meeting area of the room, where stand-up meetings occur on daily basis in front of several white boards.

The frequency response figures demonstrates pre- and post-calibration curves in the same figure with the red curve representing the frequency response before any adjustments and the green curve representing the frequency response after applying the suggested EQ settings by applying the equalization distribution theory. The EQ values each loudspeaker received after all profiles were calculated and set can be seen in Fig. 6.10 and Fig. 6.11 for calibration using white noise and calibration using the tonal signal, respectively. Blue curve represents loudspeaker A. Red curve represents loudspeaker B. Yellow curve represents loudspeaker C

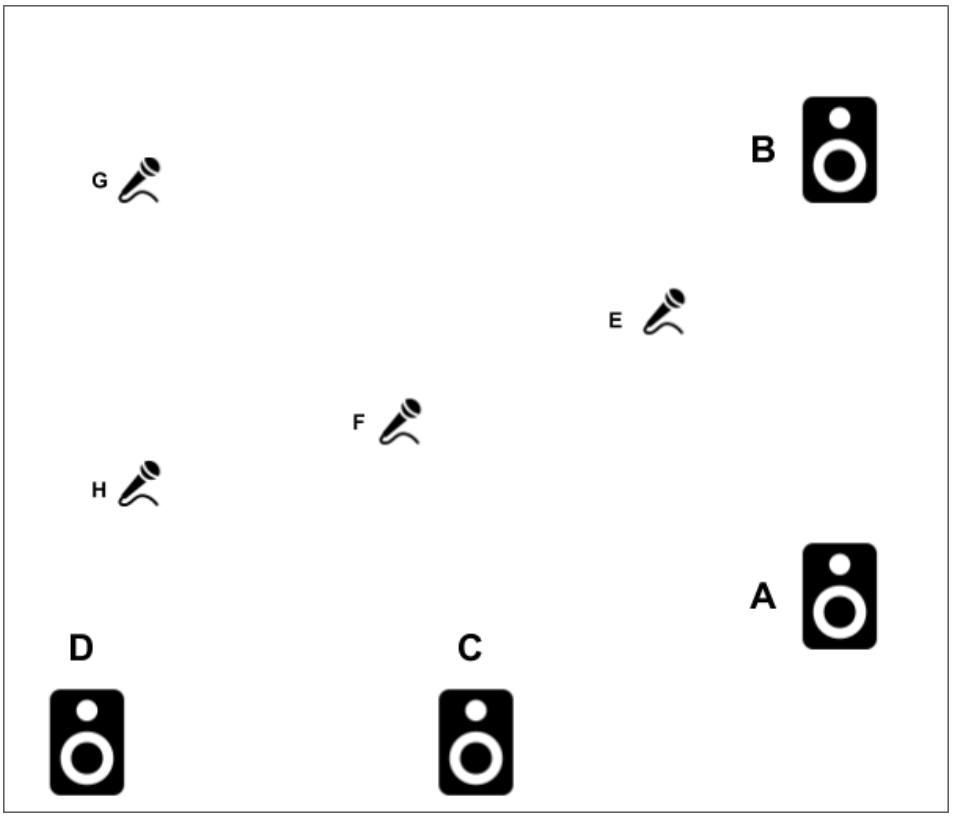


Figure 6.1: Setup of loudspeakers A-D and points of interest E-H in the reverberations room.

and purple curve represents loudspeaker D. Table 6.2 and Table 6.3 shows the SD results for PoI E-H according to setup in Fig.6.1 and the achieved gain when calibrating using white noise and the tonal signal, respectively.

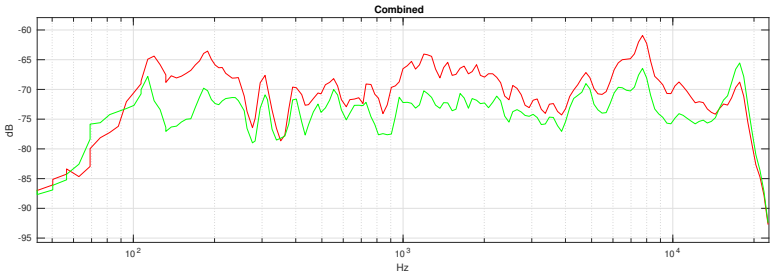


Figure 6.2: Frequency response in PoI E using noise calibration.

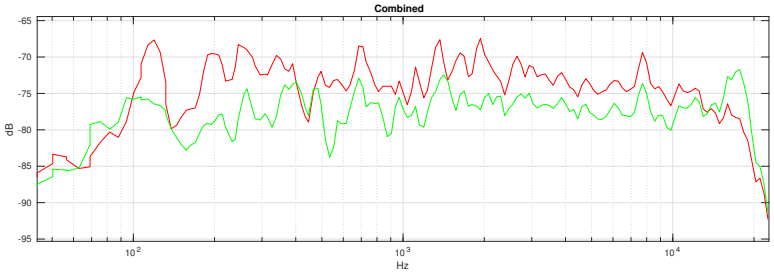


Figure 6.3: Frequency response in PoI F using noise calibration.

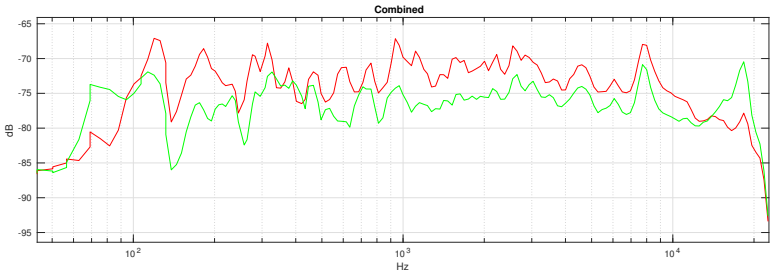


Figure 6.4: Frequency response in PoI G using noise calibration.

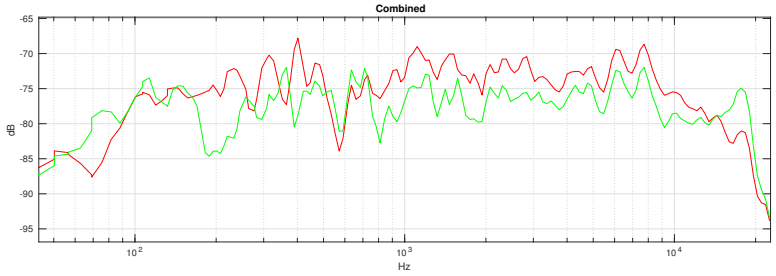


Figure 6.5: Frequency response in Pol H using noise calibration.

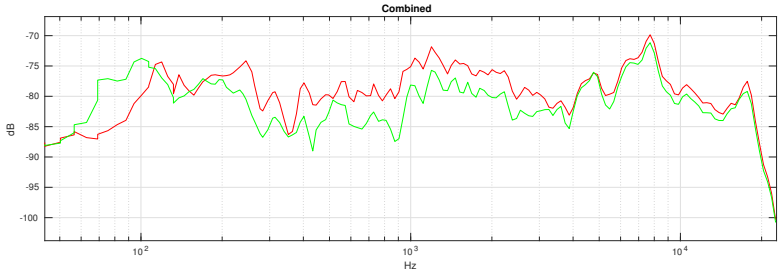


Figure 6.6: Frequency response in Pol E using tonal calibration.

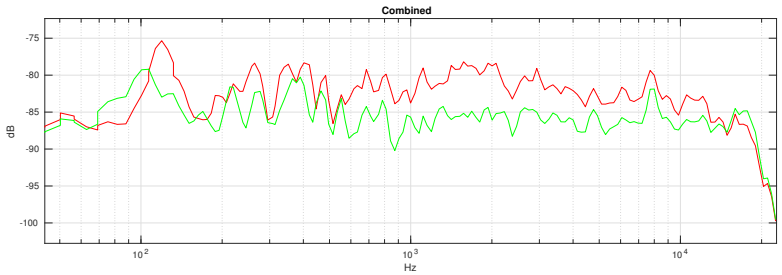


Figure 6.7: Frequency response in Pol F using tonal calibration.

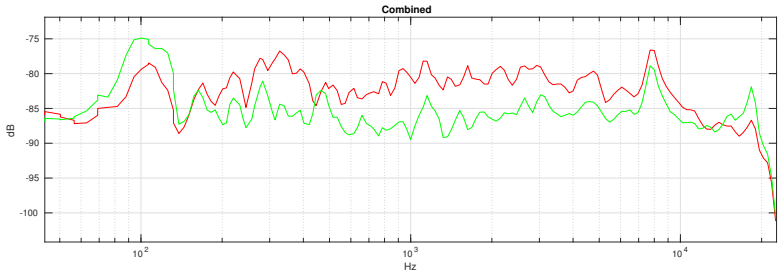


Figure 6.8: Frequency response in Pol G using tonal calibration.

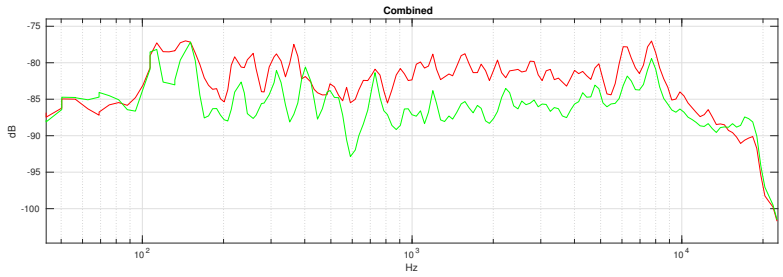


Figure 6.9: Frequency response in PoI H using tonal calibration.

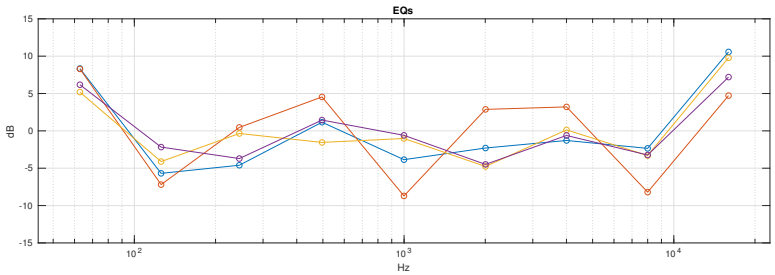


Figure 6.10: EQ values applied to each loudspeaker noise calibration where the blue, red, yellow and purple curves represents loudspeaker A, B, C and D, respectively.

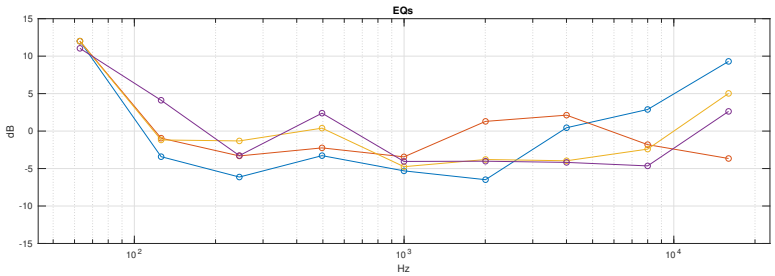


Figure 6.11: EQ values applied to each loudspeaker using tonal calibration where the blue, red, yellow and purple curves represents loudspeaker A, B, C and D, respectively.

PoI	σ_{before}	σ_{after}	G_{before}	G_{target}	G_{after}
E	3.33	1.15	-31.80	-45.00	-44.38
F	3.69	1.33	-37.13	-50.00	-47.32
G	3.66	0.73	-36.67	-45.00	-46.96
H	3.91	1.62	-37.38	-40.00	-48.51

Table 6.2: Standard deviation before and after calibration as well as current gain, anticipated gain and resulting gain when using white noise calibration.

PoI	σ_{before}	σ_{after}	G_{before}	G_{target}	G_{after}
E	2.95	2.55	-40.62	-45.00	-49.97
F	2.83	1.39	-46.01	-50.00	-55.66
G	2.86	2.51	-45.55	-45.00	-55.04
H	3.56	2.32	-46.01	-40.00	-55.82

Table 6.3: Standard deviation before and after calibration as well as current gain, anticipated gain and resulting gain when using tonal calibration.

Chapter 7
Conclusion

The localization and distance calculation method proved to work very well as it calculated the coordinates for each loudspeaker accurately as seen by the results. It can be assumed that the method assures fairly accurate positions with setups in similarity to the testing setup. However, should there be obstacles in room, for instance walls or clothing racks, there may be a minor disturbance in the signal transmitting which could affect the calculated distances and thereby distort the results since any direct waves will be blocked.

Both the single and multi channel calibration shows results where the overall sound image has improved in every PoI, showing that the equalization distribution theory works properly. Using either the tonal signal or white noise as the calibration signal proves to be more effective than not calibrating at all, as shown by the frequency response curves and SD for each octave band. Furthermore, it's evident that using white noise as calibration was superior to using the tonal signal for calibration in terms of reaching the target frequency response. However, both methods have their strengths and weaknesses in the form of effectiveness and time consumption. The white noise method proves to be more effective than the tonal signal in terms of results, whereas using the white noise method would mean a calibration where each loudspeaker in the system requires 30 s of white noise for the calibration to be accurate, which scales linearly with the number of loudspeakers. Using the tonal signal for calibration proves to be considerably faster with only requiring 3 s per loudspeaker scaling linearly. As the anechoic chamber has no reverberations or other similarities to a real environment, the frequency responses were already relatively flat before any adjustments. That means the post-calibration response was not very different from the pre-calibration response. The PoI calibration was successful as it came very close to the desired sound levels when using the specified distribution algorithm, especially while using the white noise calibration. It should therefore be plausible to suggest a desired sound level for the post-calibration.

For use in real environments, the last and first step of the calibration will not be needed in the final product as it only acted as a verification in the testing phase to confirm the feasibility of the method of generating a flat frequency response. The calibration needed for the system in real environments involves playing the 30 s white noise signal from all included loudspeakers consecutively. Afterwards, the

loudspeakers may be individually adjusted to fit a certain customer sound profile. Another aspect to take in to account is the prospect of the developed method, it could be expanded to entail even more features and functions depending on which loudspeaker it is used on, which could lead to even more optimization. An EQ containing more frequency bands than 9 will likely get a more flat result in the frequency response correction curves as it allows adjustments to more points in the spectrum. The width of the bands in the EQ used makes it hard to adjust certain frequencies without affecting nearby that you wish to remain unchanged. With more bands this could have been avoided as each band would have been more narrow which could allow peaks to be attenuated fully without attenuating more than needed on nearby frequencies.

However, using the C1004-E loudspeaker and the T8353 microphone posed several limitations. Since networked devices will not achieve perfect synchronization, this has to be taken into consideration when formulating the sound image correction method. Verifying PoI with multiple loudspeakers will generate frequency responses with different results due to clock drift and synchronization problems, especially while using tonal signals. To overcome this limitation, the method proposed uses noise as verification to minimize interference between loudspeakers. Due to poor shielding and loudspeaker output limitations, frequencies below 70 Hz could not be reliably recorded, as shown in the correction factor results. Since the results of the CF tests are close to 1, the CF was decided not to be included in the final implementation. That is, the calculations for the single and multi channel calibrations were done without the use of the CF.

7.1 Future work

The whole thesis was based upon the potential of networked loudspeakers and microphones. Passive loudspeakers connected to a central amplifier will always have better synchronization compared to networked loudspeakers, which proved to be lacking synchronization-wise. Using the assumption of having perfect synchronization of devices opens up the possibility of calibrating the whole system at the same time instead of running every device consecutively. This is something that can be investigated in future work. There could also be different ways of distributing EQ settings to different loudspeakers based on different parameters instead of using the recorded magnitude as an absolute method which could be something to investigate in the future. The results are heavily dependent on the filtering method used which leads to the conclusion that having more bands to adjust would lead to an improved result. There might also be possible to use frequency sweeps for the calibration in order to get the advantages from both the tonal signal and the white noise.

References

[1] Axis Communications AB (2013). *Frequency response*. [image] Available at: https://www.axis.com/files/datasheet/ds_t83_54493_en_1401_hi.pdf [Accessed 23 May 2018].

[2] Chung, K. and Mongeau, L. and McKibben, N. (2009). "Wind noise in hearing aids with directional and omnidirectional microphones: Polar characteristics of behind-the-ear hearing aids". *The Journal of the Acoustical Society of America*, 125(4), pp. 2243–2259.

[3] Chung, K. (2012). "Wind noise in hearing aids: II. Effect of microphone directivity". *International journal of audiology*, 51, pp. 29-42.

[4] Chung, K. (2012). "Comparisons of spectral characteristics of wind noise between omnidirectional and directional microphones". *The Journal of the Acoustical Society of America*, 131(6), pp. 4508–4517.

[5] Chung, K. (2011). "Microphone Directionality, Pre-Emphasis Filter, and Wind Noise in Cochlear Implants". *The Journal of the American Academy of Audiology*, 22(9), pp. 586-600.

[6] Chung, K. and McKibben, N. (2011). "Microphone Directionality, Pre-Emphasis Filter, and Wind Noise in Cochlear Implants". *The Journal of the Acoustical Society of America*, 22(9), pp. 586-600.

[7] Yamaha Corporation. (2018). *Microphone directionality*. [image] Available at: http://www.yamahaproaudio.com/global/en/training_support/selftraining/pa_guide_beginner/microphone/#headerArea [Accessed 23 May 2018].

[8] Massenburg, G. (1972). Parametric Equalization. In: Audio Engineering Society Convention 42. [online] Los Angeles: Audio Engineering Society, pp. 2-3. Available at: http://www.aes.org/tmpFiles/elib/20180508/16171.pdf [Accessed 6 May 2018].

[9] Bailey, D. and Swarztrauber, P. (1993). "A Fast Method for the Numerical Evaluation of Continuous Fourier and Laplace Transforms". SIAM Journal on Scientific Computing, 15(5), pp. 1105-1110.

[10] MediaCollege.com, (1995). Parametric Equalizers. [online] Available at: https://www.mediacollege.com/audio/eq/parametric.html [Accessed 6 May 2018].

[11] www.ni.com, (2010). "Characteristics of Different Smoothing Windows". [online] Available at:http://zone.ni.com/reference/en-XX/help/370051T-01/cvi/libref/analysisconcepts/characteristics_of_different_smoothing_windows [Accessed 29 April 2018].

[12] Nechyba, M. (2003). "Spectral Leakage and Windowing". [PDF] Benton: University of Florida, 5. Available at: https://mil.ufl.edu/nechyba/www/__eel3135.s2003/lectures/lecture19/spectral_leakage.pdf [Accessed 5 May 2018].

[13] Lyon, D. A. (2009). "The Discrete Fourier Transform, Part 4: Spectral Leakage". *Journal of Object Technology*, [online] 8(7), pp. 23-34.

[14] Harris, J, F. (1976). "Windows, Harmonic Analysis, and the Discrete Fourier Transform". NUC TP532, Nay. Undersea Center, San Diego, CA.

[15] Harris, J, F. (1978). "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform". *Proceedings of the IEEE*, 66(1), pp. 51-83.

[16] Welch, P. (1967). "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 Signal Processing*, 15(2), pp. 70-73.

[17] U.S. Department of Energy Office of Scientific and Technical Information, Solomon Jr, O. (1991). "PSD Computations Using Welch's Method". Springfield, VA: National Technical Information Service.

[18] www.mathworks.com, (2015). "Welch's Method". [online] Available at:https://www.mathworks.com/examples/signal/mw/signal-ex45136368-welch-s-method [Accessed 28 April 2018].

[19] Parker, G. (1991). "Using a discrete fourier transform to estimate phase in interferometric direction finding". Canberra: Australian Government Publishing Service, p.4.

[20] National Instruments, (2016). Understanding FFTs and Windowing. [online] www.ni.com. Available at: http://download.ni.com/evaluation/pxi/Understanding%20FFTs%20and%20Windowing.pdf [Accessed 2 May 2018].

[21] Smith, J.O. and Serra, X. (1985). "An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation". [pdf] Stanford: Center for Computer Research in Music and Acoustics, Standford University. pp. 1-20. Available at: https://ccrma.stanford.edu/~jos/parshl/parshl.pdf [Accessed 2 May 2018].

[22] Smith, J.O. (2011). [ebook] "Spectral Audio Signal Processing". W3K publishing. Available at: https://ccrma.stanford.edu/~jos/sasp/Welch_s_Method.html#19252 [Accessed 2 May 2018].

[23] Smith, J.O. (2011). [ebook] "Spectral Audio Signal Processing". W3K Publishing. Available at: https://ccrma.stanford.edu/~jos/sasp/Resolution_versus_Stability.html [Accessed 4 May 2018].

[24] Iwata, H., Umebayashi, K., Tiiro, S., Lehtomäki, J, J., López-Benítez, M. and Suzuki, Y. (2016). "Welch FFT Segment Size Selection for Spectrum Awareness System". *IEICE Transactions on Communications*, E99-B (8).

[25] Smith, J.O. and Serra, X. (1985). "An Analysis/Synthesis Program for Non-Harmonic Sounds Based on a Sinusoidal Representation". [pdf] Stanford: Center for Computer Research in Music and Acoustics, Standford University. pp. 6-8. Available at: https://ccrma.stanford.edu/~jos/parshl/parshl.pdf [Accessed 17 May 2018]

[26] Proakis, J. and Manolakis, D. (1996). "Digital Signal Processing, Principles, Algorithms, and Applications" (third edition) New Jersey: Prentice-Hall, Inc. pp. 480-481.

[27] MIT (Department of Electrical Engineering and Computer Science). "The Goertzel Algorithm and the Chirp Transform", (2006), Lecture 20.

[28] Sysel, P. and Rajmic, P. (2012). "Goertzel algorithm generalized to non-integer multiples of fundamental frequency". *Journal on Advances in Signal Processing* [online] 1(56), pp. 1-8. Available at: https://pdfs.semanticscholar.org/3493/dc3ccf5bdb2ffc728e4d59f5f186a70a26ca.pdf [Accessed 2 April 2018].

[29] Proakis, J. and Manolakis, D. (1996). "Digital Signal Processing, Principles, Algorithms, and Applications" (third edition) New Jersey: Prentice-Hall, Inc. pp. 477-479.

[30] Tingay, J. (2011). "Frequency Weighting Curves – 'A', 'C' & 'Z'". [image] Available at: https://www.cirrusresearch.co.uk/blog/2011/08/what-are-a-c-z-frequency-weightings/ [Accessed 7 May 2018].

[31] PJS. (2017). "Periodic sine wave without leakage (red), non-periodic sine wave with leakage (green), and windowed non-periodic sine wave with reduced leakage (blue)". [image] Available at: https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/Windows-and-Spectral-Leakage/ta-p/432760 [Accessed 2 May 2018].

[32] Neagoe, T., Cristea, V. and Banica, L. (2006). "NTP versus PTP in Computer Networks Clock Synchronization". In: IEEE International Symposium on Industrial Electronics. [online] Montreal, Canada. IEEE, pp. 338-340. Available at: https://ieeexplore.ieee.org/document/4077944/ [Accessed 8 May 2018].

[33] Annibale, P., Filos, J., Naylor, P, A. and Rabenstein R. (2012). "TDOA-Based Speed of Sound Estimation for Air Temperature and Room Geometry Inference". *IEEE Transactions on Audio, Speech, and Language Processing*, 21(2).

[34] www.siemens.com, (2016). "What is A-weighting?". [online] Available at: https://community.plm.automation.siemens.com/t5/Testing-Knowledge-Base/What-is-A-weighting/ta-p/357894 [Accessed 4 May 2018].

[35] Wang, L., Qu, C. and Yan, X. (2017). "Design of power over ethernet switchboard based on Gigabit network interface". In: Chinese Automation Congress (CAC). [online] Jinan, China: IEEE, pages. Available at: https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=8242771 [Accessed 9 May 2018].

[36] www.gstreamer.org "Gstreamer" [online] Available at: https://gstreamer.freedesktop.org/

[37] Doctor ProAudio (2012) Dynamics processors. Available at: http://www.doctorproaudio.com/doctor/temas/dynamics-processorscompressors_en.shtml [Accessed 22 May 2018].

[38] Kadis, J. (2008) Dynamic Range Processing and Digital Effects.

[39] Gan, S, W., Kuo, M, S. and Feng, J, W. (2005). "Adaptive Noise Equalizer with Equal-Loudness Compensation". In: IEEE International Symposium on Circuits and Systems [online] Kobe, Japan: IEEE. Available at: https://ieeexplore.ieee.org/document/1464578/ [Accessed 10 April 2018].

[40] Icons made by [Freepik, Simpleicon and Google] from www.flaticon.com.

[41] Watkinson, J. (1998). "The Art of Sound Reproduction". Oxford: Focal Press, pp. 139-155.

[42] Watkinson, J. (1998). "The Art of Sound Reproduction". Oxford: Focal Press, pp. 28-32.

[43] Watkinson, J. (1998). "The Art of Sound Reproduction". Oxford: Focal Press, pp. 99-102.

[44] Kuo, H-H. (1996). "White Noise Distribution Theory". Vol. 5. CRC press.

[45] Lacanette, K. (1991). "A Basic Introduction to Filters - Active, Passive, and Switched-Capacitor". Available at: http://www.swarthmore.edu/NatSci/echeeve1/Ref/DataSheet/IntroToFilters.pdf [Accessed 23 May 2018].