NTNU
Norwegian University of
Science and Technology

Optimizing Microphone Arrays for use in Conference Halls

Andreas Holm

Master of Science in Electronics
Submission date: June 2009
Supervisor: Peter Svensson, IET
Co-supervisor: Morgan Kjølerbakken, SquareHead Technology

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

Problem Description

The student will investigate the use of microphone arrays for use in a conference hall setting. The main goal is to find a way to get the sound output to be played loud enough for the audience on a public address system. When using the microphone array for pickup of sound from the audience in an auditorium, the playback can be affected by placement of array, feedback from the public address system, delay from the travel time, processing time, and the problem of getting loud enough volume so that the audience will be able to hear what is being said. These elements will be investigated along with the optimal array placement and the optimal kind of room.

Array techniques that can be used to improve these challenges will be studied, and the use of adaptive filters and the possibility to attenuate certain directions will also be investigated. In addition, there will be added filters to make the frequency spectrum completely flat or best suited for speech.

Assignment given: 15. January 2009
Supervisor: Peter Svensson, IET

Abstract

In conference halls, the question and answer session is often impaired by poor solutions for listening to the audience on the public address system. By using microphone arrays to capture the sound from questions, everyone can hear what the discussion is all about. If a tracking is used as well, these arrays creates an automatic speaker and sound reproduction environment.

Because of issues like feedback, delay, and frequency response, a feasibility study on the use of microphone arrays in conference halls has been done. By combining and applying known feedback techniques as well as newly developed methods, this study shows that a 10 dB louder gain before feedback margin has been achieved and thereby making the automatic sound capture system work. Issues still remain, and many of the feedback techniques used can linearly distort the frequency response and reduce the sound quality. Latency in the processing time has been reduced in this thesis by 20 ms, and the total delay with 15 meters sound travel time and a loudspeaker array is measured to be 58.2 ms. However, according to the Haas effect, the total delay should be below 40 ms to avoid echoes.

In this thesis, the best performing setup that gives the loudest maximum stable gain is achieved with a microphone array tilted above the second row to hit in the middle of the audience in an auditorium, loudspeaker arrays mounted high on the front wall programmed to play down towards the audience, a frequency-shifter of -5 Hz, a bandwidth from 500 to 6500 Hz, and a soft equalization along with an automatic tracker for locating the speaker. However, the sound quality difference between a close-up microphone and the microphone array is still present, and the lack of low and middle frequencies in the microphone array makes the microphone array sound a little narrow.

Preface

This thesis is part of my Master of Science degree at the Norwegian University of Science and Technology. I have worked with steerable microphone arrays in conference halls in collaboration with SquareHead Technology AS.

The measurements and tests have been performed at different locations, but for the most part at SquareHead Technology's office in Nydalen, Oslo. The tests have been designed by myself after tips from both research papers, colleagues, and supervisors.

I would like to thank Prof. Peter Svensson at the Norwegian University of Science and Technology, for being my subject teacher, and would like to thank SquareHead Technology and Morgan Kjølerbakken for providing equipment, suggesting this project, and for being my supervisor. In addition, I would like to thank Prof. Sverre Holm at the University of Oslo, for advising me in my work and Ines Hafinovic and Carl-Inge Colombo Nilsen at SquareHead Technology and the University of Oslo, for collaboration.

NTNU Trondheim, 11. June 2009
Andreas Holm

Contents

List of Figures

Chapter 1

Introduction

1.1 Task: Optimizing Microphone Arrays for use in Conference Halls

The task formulation for this master thesis is to optimize microphone arrays for use in conference halls, and this includes many areas of interests. The most important ones are:

- Feedback Suppression
- Delay
- Frequency Spectrum
- SNR and Noise Suppression
- Automatic Tracking

The main goal for this thesis, is to find a way to get the sound output from a microphone array created by SquareHead, to be played loud enough for an audience to hear clearly what is being said. When using the microphone array for capturing sound from the audience in an auditorium, the playback can be affected by placement of array, feedback from the public address system, delay from the travel time, and processing time along with the optimal loudspeaker placement and the optimal room shape. Array techniques that can be used to improve these challenges, will be studied, and the use of adaptive filters and the possibility to attenuate unwanted directions, will be investigated. In addition, filters will be added to make the frequency spectrum completely flat or best suited for speech playback.

When looking into all these considerations, an optimal system will be designed and setup in a conference hall in Stavanger (Fig. 1.1).

Figure 1.1: The AudioScope in a Conference Hall

1.2 Background

When people gather in meetings, it is common to use large conference halls to address all the participants. This will ensure the best opportunity for everyone to hear what is being said from the stage, and it is easy to capture the speaker's voice for public address playback. The conference works well for a normal lecture or presentation, but when the audience wants to interact with questions or debate contributions, the conference hall can be a troublesome setting. Transmitted sound in large conference halls is a sum of many multi-path transmissions from reflections of ceiling, floor, and walls. Since the reverberation time in large conference halls is long, the direct sound is weak compared to the diffuse sound field and noise sources can mask sound signals. People will often ask questions without being heard by the other participants which can be frustrating for the audience.

To maximize the ratio of direct sound to reverberant sound, which will provide the strongest and cleanest signal, it is important to have a microphone as close as possible to the sound source. However, in large conference halls, this can be problematic since many microphones needs to be placed either near every participant or a microphone has to be passed around to people who wants to interact. A mixer is also needed to add all the microphone signals together which increases the technical complexity. Since one microphone is needed at every seat, it can often be in the way of papers or view. The pass around solution is the most common today, and it makes the interactive discussion very slow and difficult because of the passing delay. This creates a high threshold for talking in meetings, and many good discussions or questions might never find place. In many smaller meeting rooms, microphones mounted from the ceiling hanging down towards the audience have been used. This solution hides the microphones, but the microphones are far from the audience and the audio can sound reverberant and cause feedback [1]. Feedback is experienced when the open loop gain between the microphone and the loudspeaker exceeds 1.

Highly directional microphones can be used to remove some of these known issues by reducing the reverberant sound and noise. Directional shotgun microphones have been made, but they have a fixed directivity and need to be handheld, and because of the directivity of 60 degrees at 2 kHz, they will have a maximum distance. Steerable microphone arrays have been researched as a problem solver for many years, but have until today suffered from the complexity of the number of microphones needed and the processing power required. Medical ultrasound machines and sonars that uses the same technique have solved this issue by making hardwired solutions for beamforming, but this is an expensive method and has been too expensive for conference halls. However, in the later years, processing power has increased above the limit that is required for software beamforming and digital microphones have removed the A/D converter requirement. This means that large microphone arrays can now be processed to create a reasonable audio output at a fair price.

Several small microphone arrays have been created for teleconference pickup and similar small applications, like the Voice Tracker by Acoustic Magic[1], but they are too small to perform satisfactorily in large conference halls. Flanagan has among others been a pioneer in making two dimensional arrays for meeting rooms, and he has researched this for some time. In [2] and [3], he investigates the use of two dimensional arrays for sound capturing in conference and discusses the bandwidth requirements. In [4], the use of circular arrays in meetings have been investigated, and in later years, MIT[2] has made a 1020 elements microphone array [5], but this array has not been used outside the laboratories.

1.3 Feedback Suppression

Even though the microphone array is highly directional, feedback will occur when the microphone output is gained loud enough. The diffuse sound field will eventually have a too high sound pressure level that feedback is difficult to avoid. The feedback phenomena has been around since Fletcher first described it in 1926 [6], and ever since then, methods to overcome this issue have been researched.

The feedback is experienced in live microphone situations like a public address system, speech meetings, tele- and video conferences, and in hearing aid equipment [7]. In reverberant rooms, feedback will occur easier than in

[1] http://www.acousticmagic.com/

[2] Massachusetts Institute of Technology

less reverberant rooms because of the large reflection surfaces and the long delay time.

Conventional methods for suppressing feedback have been developed for different applications than microphone arrays, but some of them might prove effective. The most known anti-feedback method is to use notch-filters [8] at the howling frequencies and thereby getting an overall larger gain for the other frequencies. Another method is to shift the whole band upward or downward by a small modulation and thereby reduce the chance for feedback. This method is called frequency-shifting [9], and the advantages of both automatic notching and frequency-shifting can be useful.

Feedback suppression with a predictor and adaptive filtering has been formerly researched with good results in [10] and [11]. A theoretical gain of 10 dB has been found and can be essential to get the microphone arrays gain sufficient enough for playback in reverberant rooms [12]. However, the filtering might linearly distort the signal so it can be too audible, and a possible delay from the algorithm can make the adaptive filtering not usable.

When looking at the feedback issue, there can be an advantage to use digital signal processing to nullify unwanted directions that causes much feedback energy [13]. This can be performed by creating a spatial filter and thereby nullify the loudspeaker source or a large reflection.

1.4 Delay

When picking up sound from a far distance, the travel time from the source to the microphone will make a delay for the sound output. In addition to this travel delay, the processing delay of the microphone array will be added to the output. This delay can be heard as annoying echoes if it is larger than 30-40 ms, but if the delay is below 40 ms, the Haas effect [14] will ensure that the sound output is echo-free. The delay needs to be kept at the minimum to avoid too annoying echoes and to reduce the reverberant sound level in a room for minimizing feedback [1].

1.5 Frequency Spectrum

When listening to audio from recordings, it is important that the recording sounds like it would be in real life and not linearly distorted. To ensure that the signal is unchanged, it is important to make the frequency spectrum of the microphone array flat which will make sure the signal is not colored.

The optimal placement of both microphone array and loudspeakers needs to be found since this can affect both the frequency response and feedback stability. What kind of loudspeaker to use and the possibilities of using loudspeaker arrays for playback, might also provide a better frequency response.

1.6 SNR and Noise Suppression

To be able to use the microphone array optimally, noise must be suppressed. If there is a lot of noise in the recorded signal, it will be harder to separate the signal from the noise, and feedback, which is highly unwanted, will occur. If there is high SNR, it will be much easier to make a good sound reproduction through the microphone array. The noise can come from electrical design, but also from signal processing steps, and by reducing this noise, the best performing setup can be found.

1.7 Automatic Tracking

An investigation on the use of tracking to steer the microphone array, will be performed. This has been investigated for many years [15], but there has been trouble controlling it, and a well-functioning tracker will be the breakthrough for this application. In [16], a microphone array that recognize a speaker is created and might provide useful information for the developed tracker.

1.8 This Thesis

This thesis begins with a theoretical array description in chapter 2 and continues with an analysis of the conference hall in chapter 3. From these two chapters, several tests have been suggested and they are performed throughout chapter 4. The analysis of the different test results is further given in chapter 5, before a conclusion from the whole investigation is written in chapter 6. Since many problems have appeared throughout this study, a future work part is summed up in chapter 6.1.

Chapter 2

Array Theory

Array theory has through time been developed so several measurement sources combined can create better detection. One example is that temperature sensors at different locations in a land can be used to create measurements. With sampling every hour, this can be a great tool for analyzing trends in the weather and to predict future weather. Array of sources can create new dimensions to measurement techniques and have many different applications today.

The use of microphone arrays is correlated with the processing power and hence the cost of the technique. Since every microphone channel needs an A/D converter and transmission to the processing station, the number of microphone channels is limited. Today, large microphone arrays are used in high cost applications such as ultrasound, sonar, and seismic, and these applications have mostly used water as a medium and not air. Microphone arrays have been created for sound capturing in teleconference setups, but they are small and do not provide the sharpest directivity. Noise tracking is also an application that uses microphone arrays.

The directivity of an array microphone is a large advantage and is often much sharper than a normal cardioid response. A normal microphone has to be physically turned to steer the direction of the sound capturing, and this means that either a person has to hold the microphone or it has to be mechanically steered to change the direction of the directivity. In a microphone array, the directivity is electronically steered, and it can be steered to any direction in front of the array without moving the array. This can be done by selecting a direction and changing the delay and weighting to each microphone element so audio from the desired direction is amplified.

The combination of several receivers or senders can result in electronically steerable high directivity beamformers. These beams can be used to separate wanted signal from surrounding noise and unwanted signals. Together with fast A/D converters and data storing, beamformers can be used to create real-time sound applications.

For a more in-depth discussion of key microphone array processing techniques, the interested reader is referred to [17].

2.1 Wave Theory

In "Two Decades of Array Signal Processing Research" by Krim and Viberg [18], a complete investigation of array theory has been done and many well known array signal processing algorithms have been compared. This is a very thoroughly written paper and studies the basics of array theory.

Many physical phenomena result from displacement of molecules or exhibition of wave-like physical manifestation. A wave propagation may take various forms, both as electromagnetic waves and acoustical waves which both follows from the homogenous solution of the wave equation. The electromagnetic wave equation can be derived from the Maxwell equations, but since the acoustical wave is not a function of electric and magnetic field, these cannot be used for deriving the acoustical wave equation. Instead Newton's laws can be used to start with.

First a few variables need to be defined.

$$\begin{aligned}
\rho &= mass\ per\ unit\ of\ the\ fluid \\
u &= velocity\ flow\ of\ fluid\ in\ x - direction \\
v &= velocity\ flow\ of\ fluid\ in\ y - direction \\
w &= velocity\ flow\ of\ fluid\ in\ z - direction \\
P &= pressure\ in\ fluid
\end{aligned}$$

Newton's law of momentum conservation says that a small volume within a gas will accelerate if there is an applied force. The force arises from pressure differences at opposite sides of the small volume. Newton's law says that

$$\text{mass} \times \text{acceleration} = \text{force} = -\text{pressure gradient} \tag{2.1}$$

$$\rho \frac{\partial u}{\partial t} = -\frac{\partial P}{\partial x} \tag{2.2}$$

$$\rho \frac{\partial v}{\partial t} = -\frac{\partial P}{\partial y} \tag{2.3}$$

$$\rho \frac{\partial w}{\partial t} = -\frac{\partial P}{\partial z} \tag{2.4}$$

The second physical process is energy storage by compression and volume change which is called Hooke's law. If the velocity vector u at $x + \Delta x$ exceeds that at x, then the flow is said to be diverging. In other words, the small volume between x and $x + \Delta x$ is expanding. This expansion must lead to a pressure drop. The amount of the pressure drop is in proportion to a property of the fluid called its incompressibility K. In one dimension, the equation is

$$\text{pressure drop} = (\text{incompressibility}) \times (\text{divergence of velocity}) \tag{2.5}$$

$$-\frac{\partial P}{\partial t} = K\frac{\partial u}{\partial x} \tag{2.6}$$

In three dimensions, it is

$$-\frac{\partial P}{\partial t} = K(\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z}). \tag{2.7}$$

To derive the one-dimensional wave equation from Eq. 2.2 and Eq. 2.6, divide Eq. 2.2 by ρ and take the x-derivate.

$$\frac{\partial}{\partial x}\frac{\partial u}{\partial t} = -\frac{\partial}{\partial x}\frac{1}{\rho}\frac{\partial P}{\partial x} \tag{2.8}$$

Then take the time-derivate of Eq. 2.6 where K is assumed to be constant.

$$\frac{\partial^2 P}{\partial t^2} = -K\frac{\partial}{\partial t}\frac{\partial u}{\partial x} \tag{2.9}$$

Finally, by inserting Eq. 2.8 into Eq. 2.9, the one-dimensional wave equation can be derived.

$$\frac{\partial^2 P}{\partial t^2} = K\frac{\partial}{\partial x}\frac{1}{\rho}\frac{\partial P}{\partial x} \tag{2.10}$$

In three dimensions and with the assumption that the wave is observed in a homogenous medium which means that $c^2 = \frac{K}{\rho}$, the wave equation can be written.

$$\frac{\partial^2 P}{\partial t^2} = c^2(\frac{\partial^2 u}{\partial x^2} + \frac{\partial^2 v}{\partial y^2} + \frac{\partial^2 w}{\partial z^2}) \tag{2.11}$$

The acoustic wave propagation in a medium will follow this equation, and in the one-dimensional case, the wave will have one solution going in positive x-direction and one solution going in the negative x-direction. The general solution is given in Eq. 2.12.

$$P = f(x - ct) + g(x + ct) \tag{2.12}$$

2.1.1 Noise Fields

There are two main noise fields that needs defining for microphone array applications. These noise fields are defined as the degree of correlation between noise signals at different spatial positions. The correlation, also called *coherence*, can be found from statistics and is defined as

$$R_{ij}(f) := \frac{\varphi_{ij}(f)}{\sqrt{\varphi_{ii}(f)\varphi_{jj}(f)}}, \tag{2.13}$$

where $\varphi_{ij}(f)$ is defined as the cross-spectral density between signal i and j. The coherence from Eq. 2.13 is bound between $0 \leq |R_{ij}(f)|^2 \leq 1$.

Coherent Noise Fields

A coherent noise field is experienced in a room where there are no reflections, dispersions, or dissipations caused by acoustical environments. This means that all sound propagates from the source directly to the microphone and creates a coherence almost equal to 1. This can be experienced in an anechoic chamber or in an open air environment where there are no reflections or wind turbulence affecting the signal.

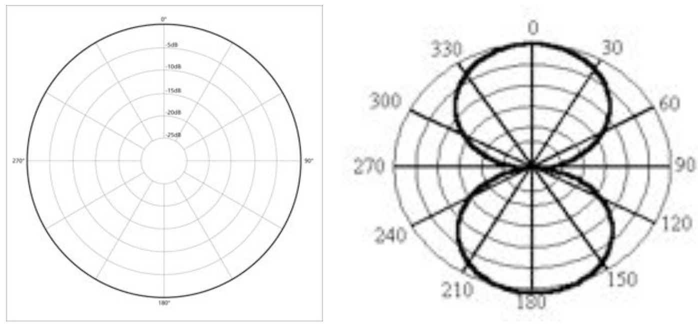
Incoherent Noise Fields

An incoherent noise field means that the noise measured at one spatial location is uncorrelated with noise measured at other locations. This means that the coherence is almost equal to 0, and there is white noise everywhere.

2.2 Directivity of Microphones

A normal microphone without any baffle will be small compared with audible wavelengths and have an omnidirectional directivity. There will be small deviations and it might not be completely omnidirectional in the high frequency response, but for array applications, this is considered good enough. This is the basic of all microphones and is the building-block for more complex microphone directivites. The omnidirectional response can be studied in Fig. 2.1(a).

When combining several microphones, directivity can be shaped by the microphone placement. The simplest array technique is based on two microphones and is called a dipole microphone. This is made by having two microphones close to each other and adding the signals together with one of the microphones with opposite phase. The directivity of a dipole microphone is shaped like a figure of eight and is a very commonly used microphone in sound studios all over the world today. When these two signals are added together, there will be destructive interference exactly to the side of those two microphones. However, in the straight line direction, there will be constructive interference so the response looks like a figure of eight when drawn. See Fig. 2.1(b) for an illustration.



(a) Omnidirectional (b) Figure of Eight

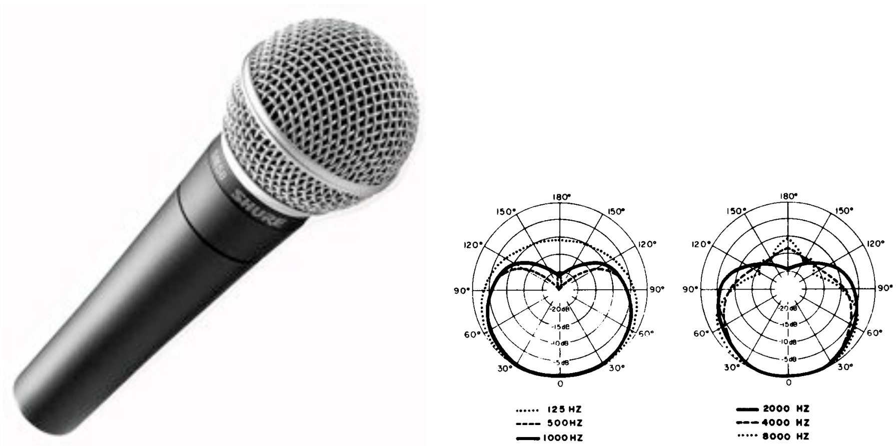
Figure 2.1: Omnidirectional and figure of eight frequency response

These microphone responses are popular in many recording studios, but the most used microphone directivity, is the combination of these two responses. The resulting response is cardioid shaped and is the most common directivity for live performance because of low amplitude response from the back of the microphone. The Shure SM58, shown in Fig. 2.2(a), is one of the most common live performance cardioid microphones. This directivity is achieved by adding omnidirectional signal with a bi-directional signal which gives the formula $\rho = (1 + cos(\theta))$. Fig. 2.2(b) shows the directivity plotted in polar coordinates.

In many situation, this directivity is good enough and no sharper directivity than this is needed. However, this is a close up microphone, and if the microphone is used further away from the source, this directivity can create feedback on a stage. By different combinations of microphone baffles, the directivity of a microphone can be much sharper than the cardioid response of the SM58, and the creation of acoustical passive arrays has proven to be very effective to create a sharp directivity. Interference in one special direction can be created by having different baffles where sound can travel to the microphone and make a sharp directivity.

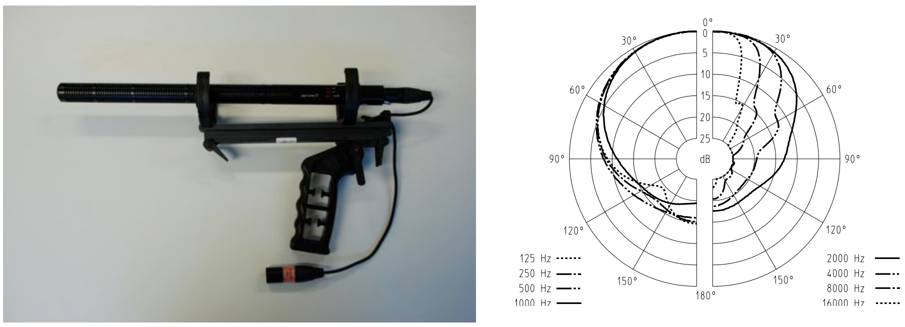
One microphone that use an acoustical array effect, is the Sennheiser MKH70 which is shown in Fig. 2.3(a). The response, shown in Fig. 2.3(b), is sharper in directivity than the SM58 in the frequencies above 2 kHz. This microphone type is called a shotgun microphone and is often used for outdoor recordings where sound pickup from distance is required. They can often be seen used to capture sound in front of a goal in a football stadium.

In all the microphones mentioned, it is physical movement of the microphone that will change the steering direction of the microphone, and this



(a) Shure SM58 (b) Cardioid Response $\rho = (1 + cos(\theta))$

Figure 2.2: Shure SM58 live microphone. Frequency response is taken from datasheet of the Shure SM58 shown in Fig. A.1.



(a) Sennheiser MKH70 (b) Shotgun Response

Figure 2.3: Sennheiser MKH70 shotgun microphone. Frequency response is taken from the Sennheiser MKH70 datasheet shown in Fig. A.2.

means that either a person has to move the microphone or some other form of mechanical movement is required. The directivity that these microphones provide is also a result of the spatial frequency response of the microphones and the baffle placement, which means that nearly any kind of directivity can be achieved by just placing microphones or acoustical passive baffles at certain places and adjusting phase and amplitude. This is what digital microphone arrays take advantage of when creating their directivity, and when many microphones are used, an array is created.

2.3 Array Aperture

Aperture is referred to a spatial region that transmits or receives propagating waves. A transmitting aperture is an active aperture, while a receiving aperture is a passive aperture. For light emission, an aperture is the size of a hole in a camera, and the antenna is the aperture in electromagnetic fields. In acoustics, this aperture is an electroacoustic transducer which is called a microphone (acoustic to electrical) or a loudspeaker (electrical to acoustical).

If a linear array is studied, the aperture size will depend on the incoming angle of the wave front, and the aperture will reduce in size when the incoming wave has a greater angle. An illustration of incoming angle and the resulting aperture size can be studied in Fig. 2.4, and the main-lobe width is inversely proportional with the aperture size.

When operating with continuous aperture given by $a(t, \mathbf{r})$ and a signal given by $x(t, \mathbf{r})$, where $\mathbf{r}$ is the spatial location in a volume and t is the time, the apertures linear filtering will be given by the convolution,

$$x_R(t, \mathbf{r}) = \int_{-\infty}^{\infty} x(\tau, \mathbf{r}) a(t - \tau, \mathbf{r}) d\tau \tag{2.14}$$

or, by taking the Fourier transform,

$$X_R(f, \mathbf{r}) = X(f, \mathbf{r}) A(f, \mathbf{r}). \tag{2.15}$$

Here, $A(f, \mathbf{r})$ is called the aperture function and describes the weighting over the aperture.

2.4 Directivity Pattern

The Fourier transform over the aperture function $a(t, \mathbf{r})$ shows the directivity pattern of the aperture. The directivity pattern is a function of frequency and the incoming angle, and the far field solution is shown in Eq. 2.16.

$$D_R(f, \theta) = \mathcal{F}(a(t, \mathbf{r})) \tag{2.16}$$

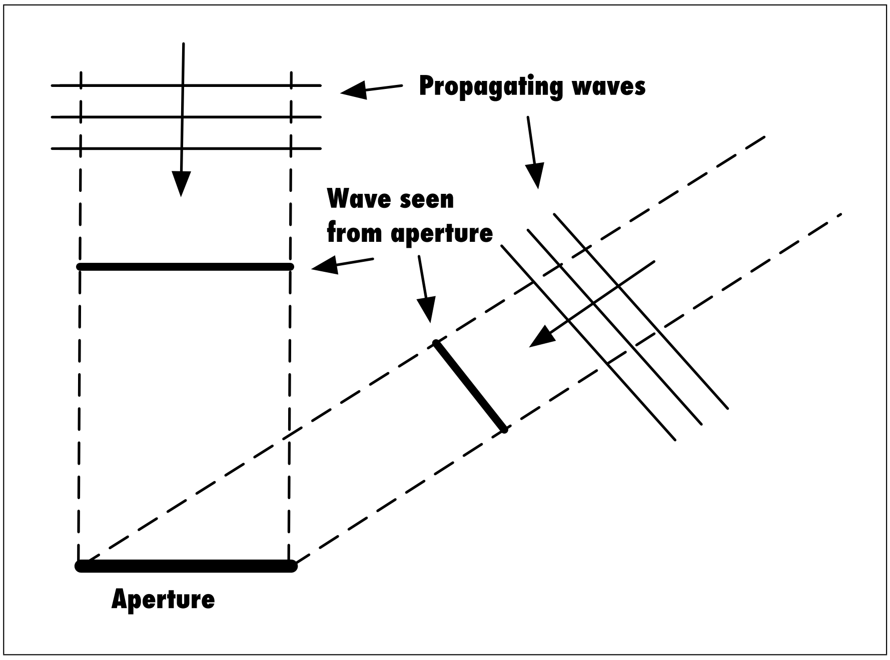


Figure 2.4: Aperture size related to incoming angle of an array.

For a straight, linear aperture, the directivity pattern in one dimension will be the Fourier transform of a rectangle function which result in a sinc function. The aperture shape affects the directivity and is one of the most important parts of a microphone array. In Fig. 2.5, the relationship between the aperture shape and the directivity is shown, and x_a is the position on the aperture, L is the length of the aperture, and α_x is the relation between wavelength and the incoming angle.

What can be seen from Fig. 2.5, is that the first zero is at λ/L, and this represent the size of the main-lobe. If the same length of array is used for all frequencies, the main-lobe width for $L = 1.55m$ will vary from 120 degrees at 200 Hz to 2.5 degrees at 10 kHz. This is a very large difference in main-lobe width and is unpractical in many situation. The aperture needs to be changed to have same main-lobe width for all frequencies which results in sub-array beamforming.

2.5 Discrete Array

Microphone arrays that are created today, are not made of a continuous aperture, but created by mounting several small microphones together. This means that the microphones will be built up of discrete sources with the distance between them as spatial sampling of the sound level at one exact position to one certain time.

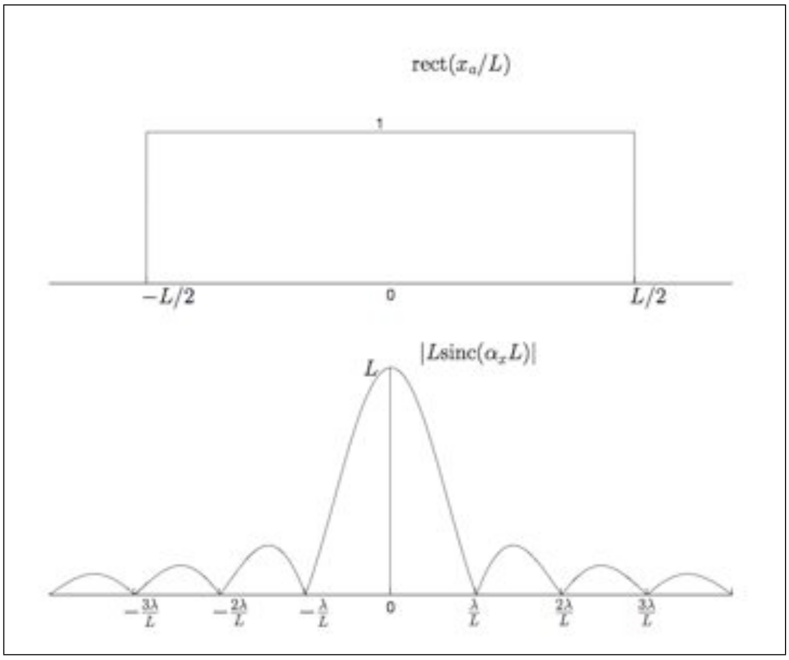


Figure 2.5: A rectangle aperture results in a sinc directivity response.

If a linear one-dimensional array is studied in one direction, with the same frequency response from all microphones and uniformly spaced placement, the relation between directivity and array is

$$D(f,\phi) = \sum_{-\frac{N-1}{2}}^{\frac{N-1}{2}} w_n(f) e^{j\frac{2\pi f}{c}\cos\phi}, \tag{2.17}$$

where ϕ is the angle and $w_n(f)$ is the complex weight function. In addition to these two variables, the directivity of the array is dependent on,

- The number of array elements, N
- The inter-element spacing, d
- The frequency, f
- The weighting, w

The side-lobe level will be reduced if the array has the same size with higher density of elements. If the size of the array is made larger by increasing the inter-element spacing, the response will show a sharper main-lobe.

2.5.1 Spatial Aliasing

As in temporal sampling, the spatial sampling of elements must be within the Nyquist limit to avoid aliasing. The temporal sampling theorem states that a signal must be sampled at a rate f_s such that

$$f_s = \frac{1}{T_s} > 2f_{max}, \tag{2.18}$$

where f_{max} is the largest frequency in the signal. Converted to spatial sampling this will be

$$f_{x_s} = \frac{1}{d} > 2f_{max}, \tag{2.19}$$

where f_{x_s} is the spatial sampling frequency and f_{max} the largest frequency that needs spatial sampling. Since $f_{max} = \frac{1}{\lambda_{min}}$ when converting the frequency into wavelength, the spatial sampling must be within the requirement

$$d < \frac{\lambda_{min}}{2}, \tag{2.20}$$

where λ_{min} is the smallest wavelength in the signal of interest. Eq. 2.20 is known as the spatial sampling theorem and must be adhered to in order to prevent the occurrence of spatial aliasing in the directivity pattern of a sensor array. Spatial aliasing will show as grating lobes in a directivity plot of a microphone array.

2.6 Beamforming

In Eq. 2.17, the relationship between the directivity and the elements spacing is mathematically shown. Until now, the $w_n(f)$ has been considered to be equal to $\frac{1}{N}$ and will instead be considered to be

$$w_n(f) = a_n e^{j\varphi_n(f)}, \tag{2.21}$$

where a_n is the amplitude weighting and $\varphi_n(f)$ is the phase weighting. The amplitude weighting will change the main-lobe width and side-lobe levels while the phase weighting will shift the main-lobe to another angle of incidence. The phase weighting is what gives the digital steering and changes the capture direction of a microphone array without physically moving the array.

A rectangular uniform amplitude weighting ($a_n = \frac{1}{N}$) will give a sinc response and thereby the sharpest main-lobe possible, but this response has high side-lobe levels, and by making a different weighting, the side-lobe level can be reduced at the expense of an increased main-lobe width. There are many known amplitude weightings and the most common is the hamming weighting. Instead of having 13.5 dB attenuation, as the uniform weighting has, the side-lobe level has a 37.2 dB attenuation with hamming weighting. The main-lobe is, however, doubled in width, but it is in many cases better to have a wide main-lobe with low side-lobes than the opposite.

There are two main methods of designing an array, and these are broadside and endfire configurations. The broadside configuration means to have microphones placed horizontally along the x-axis (Fig. 2.6), achieving the main-lobe in the y direction, while the endfire configuration means to have the microphones placed vertically along the y-axis, achieving the main-lobe also in the y direction.

Beamforming is the combination of amplitude and phase weighting, and there are many different techniques for performing this. Below is a short summary of the most common methods.

2.6.1 Delay-and-Sum Beamforming

The simplest form of beamforming is the delay-and-sum beamforming. If a linear array is created and is aligned like Fig. 2.6, the steering direction can be changed by using a complex weighting function. The delay-and-sum beamforming is created with a phase weighting like

$$\varphi_n(f) = -\frac{2\pi f(n-1)d\cos(\phi)}{c}. \tag{2.22}$$

If this phase weighting is combined with Eq. 2.21 and an amplitude weighting like $a_n = \frac{1}{N}$, then $w_n(f)$ will be

$$w_n(f) = \frac{1}{N}e^{\frac{-j2\pi f}{c}d(n-1)\cos(\phi)}. \tag{2.23}$$

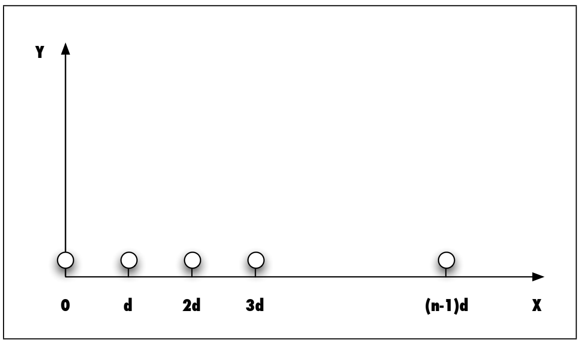


Figure 2.6: Linear one-dimensional discrete array

The output from the array can then be added together by applying the weighting function to all input channels. If the array output is expressed as the sum of the weighted channels, Eq. 2.24 can be obtained.

$$y(f) = \frac{1}{N}\sum_{n=1}^{N} x_n(f) e^{\frac{-j2\pi f}{c} d(n-1)\cos(\phi)} \tag{2.24}$$

This is the frequency domain relation, and in the time domain, it will be

$$y(t) = \frac{1}{N}\sum_{n=1}^{N} x_n(t - \frac{(n-1)d\cos(\phi)}{c}). \tag{2.25}$$

The time domain delay is the differential travel time between the reference sensor and the n^{th} sensor. When this delay is added to each element, the main-lobe can be steered to a chosen angle of incident.

The delay-and-sum beamformer needs a large array size to be able to create directivity in the low frequency region. This means that the size of the array needs to be too large to have low frequencies and is inconvenient to handle.

2.6.2 Filter-Sum Beamforming

The filter-sum beamformer is a generalized version of the delay-and-sum beamforming. Filter-sum beamforming means that both the amplitude and phase weightings is dependent on the frequency and thereby includes most of the beamformers that are used today. The output of a filter-sum beamforming is given as

$$y(f) = \sum_{n=1}^{N} w_n(f) x_n(f). \tag{2.26}$$

The $w_n(f)$ denotes the weighting related to frequency, and the $x_n(f)$ signal is divided into each frequency component. These are multiplied and the frequency weighting is performed.

2.6.3 Sub-Array Beamforming

Sub-arrays are one way of processing separate frequency bands independently to create the same main-lobe width for a wide-band signal [19]. If no sub-arrays are used, a large difference will be experienced between the main-lobe width of the highest and lowest frequencies. In an uniform linear array with i.e. 40 elements with 4 cm distance, the 1000 Hz signal will have a main-lobe of approximately 24 degrees, and a 4000 Hz main-lobe will be four times smaller with 6 degrees. If the 4000 Hz signal only use one fourth of the aperture, the main-lobe will have the same beam-width for both 1000 Hz signal and 4000 Hz signal. This can be performed for certain frequency band in the complete band-width and thereby create one main-lobe width for all frequencies.

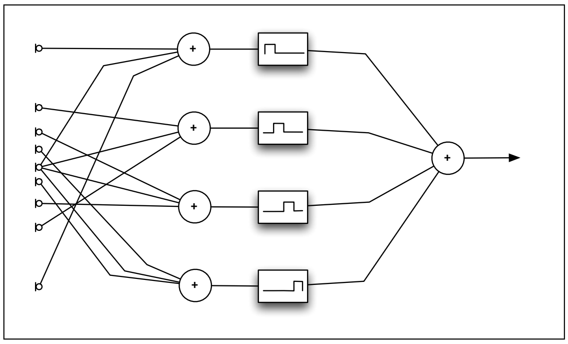


Figure 2.7: Sub-array beamforming example

Another advantage by using sub-arrays, is that the microphone placement in the array does not need to be uniform, and the density of microphones can change across the array. The sub-array processing of the lowest frequencies needs the microphones furthest from the center, and the density can increase when approaching the center of the array. The highest sub-array requires only the center microphones to get the same beam-width as the lowest sub-array. In Fig. 2.7, a one-dimensional illustration of the sub-array technique is described with 9 microphones.

2.6.4 Superdirectivity and Near Field Superdirectivity

Superdirectivity and near field superdirectivity are endfire array techniques that can create good directivity in the low frequency region by using fewer microphones and smaller size of the whole array. These techniques are based on differential adding of microphone signals and create the directivity like a dipole. However, this technique requires microphones in the steering direction to create a sharp response, and in a flat array, this will only be possible in the horizontal direction and not in the vertical direction, which is the sum-and-delay steering direction. There have been experimental methods by combining a sum of a sum-and-delay array and a superdirectivity array which have shown good theoretical performance, but they are complex to create and the array need to be three dimensional to use both techniques. [20, 21] are examples of these superdirectivity arrays, and their responses are shown in these papers.

2.6.5 Overview and Comparing of Methods

In Table 2.1, the different beamforming techniques are summarized and the advantages and disadvantages are listed. The delay-and-sum and the sub-array beamforming requires incoherent noise and a broadside array configuration, while the superdirectivity methods needs a diffuse noise field with an endfire array configuration. The endfire and broadside method of creating arrays are quite opposite and often one of these configurations is used and not both of them, but if a spherical array is created, both the superdirectivity and the sub-array technique can be combined to create a broad-band array with good low-frequency performance.

These noise conditions are seldom completely fulfilled so the theoretical SNR will be higher than any practical SNR.

Technique	Advantages	Disadvantages
Delay-and-sum	Simplicity	Low frequency performance, Narrow band
Sub-array delay-sum	Broad-band	Low frequency performance
Superdirectivity	Optimized array gain	Assume diffuse field
Near field superdirectivity	Optimized array gain, Low frequency performance	Assume diffuse field, Assume noise in far field

Table 2.1: Advantages and disadvantages of different array techniques.

2.7 SquareHead's Array Configuration

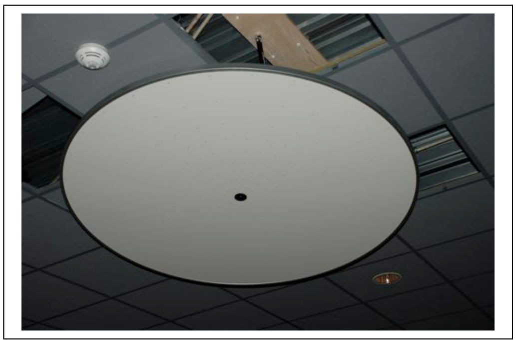


Figure 2.8: SquareHead's microphone array

The SquareHead microphone array in Fig. 2.8, is a broadside circular array called the AudioScope. It has a diameter of 1.55 meters and has 285 microphones placed on the surface. The beamforming is processed through a sub-array beamformer and provides an even main-lobe response from 800 to 5000 Hz of approximately 20 degrees. The microphones are mounted in rings which close in towards the center of the array. In the middle of the microphone array, a camera is placed to get a overview picture to the workstation. All microphone data along with the camera picture is passed to the workstation where a touchscreen can be used to steer the array's virtual microphone to the wanted listening position. The microphone used in the microphone array is the Pulse TC100E. This is a digital microphone that provides a sigma-delta modulated data stream which is sent to the workstation. The microphone array used is described in [22] and [23], and other papers about this array are [24] and [25].

Chapter 3

Conference Issues and Theories

Figure 3.1: Conference hall with an AudioScope installed

Microphone arrays have been researched for solving the conference troubles, and by placing microphones closely spaced together (below $\frac{\lambda}{2}$), a digitally controlled directional microphone can be created. In Fig. 3.1, a typical conference hall is shown with a microphone array in the ceiling. Since the directivity achieved from the array microphone is very sharp, the captured audio does not include as much reverberation from the room as other microphones have. The microphone array can be used for automatic tracking of sound sources, where the operator only needs to turn the microphone array on to amplify persons among the audience.

The explained solution sounds very simple, but there are many acoustical issues that need investigation and will be addressed in this thesis. They are:

- Feedback
- Delay
- Loudness level
- Frequency Spectrum
- Placement of Microphone array and Loudspeakers
- Automatic Tracking

3.1 Feedback

The largest acoustical issue that will occur is feedback. The whole system will go into feedback if the open loop gain, shown in Fig. 3.2, between the loudspeaker and the microphone is above 1. If

- F = Transfer function of the acoustical feedback path
- V = Amplified system output
- S = Signal source
- G = Gain

$$\frac{V(\Omega)}{S(\Omega)} = \frac{G}{1 - GF(\Omega)} \tag{3.1}$$

Feedback will occur if

$$|GF(\Omega)| > 1. \tag{3.2}$$

This will happen at different frequencies depending on phase, distance, and microphone array steering direction. There will also be large differences between rooms because a large reverberant room will easier go into feedback than a small, dampened room. The feedback occurs often because of loud diffuse sound field, and this will in many cases be the limiting factor. To reduce the diffuse sound level, the loudspeakers have to be turned down and this can result in a too low a level for the audience to hear what is being said. Techniques are therefore required to increase the gain before feedback level so the maximum stable gain is loud enough.

In both public address and video conferencing systems, the feedback issue has been present, and much research effort has been put down to improve the robustness of these setups. However, when using a microphone array, which is a new type of microphone, the old conventional anti-feedback methods needs to be reinvestigated to see whether they still can improve the gain before feedback. This means that combinations of several anti-feedback

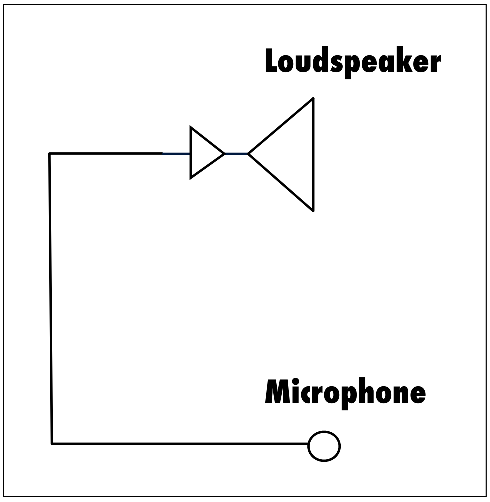


Figure 3.2: Normal feedback loop in a public address system

methods can be the optimal solution, but since some of these techniques are based on nonlinear processes, a combination of these methods can make the signal audible distorted and too low fidelity sounding.

3.1.1 Frequency-Shifter

Since the feedback issue has been around a long time, it is a well studied subject, and frequency-shifting has proven to be one of the most effective methods for suppressing feedback. The pioneers are Schroeder [9, 26] and Prestigiacomo/Maclean [27], and they first researched the method around 1960 where they discovered the possibility to shift the whole frequency band by a few Hz up or down to improve the gain before feedback. They have made the first versions of the frequency-shifter before other people have started to improve the design. Burkhard [28] and Hartley Jones [29] have created analog designs, and there have later been created digital designs for computer implementations.

In video- and teleconferencing the frequency-shifting is a very common method and is used in many anti-feedback products today. The method has proven to be very useful when controlling feedback in meetings. It can be found in telephone conferences software such as Skype[1] where it is used along with echo cancellation to achieve feedback stability.

The concept of the frequency-shifter is not a very complex idea, and it consists of having a different signal played out by the loudspeaker than what the microphone picks up. This will increase the feedback stability, and the frequency-shift should be performed on the microphone signal before it is amplified by the loudspeaker.

[1] www.skype.com

3.1.2 Notching and Equalization

An old anti-feedback method is the equalization and filtering of the frequency band to reduce feedback. This is a simple method, and equalizers are found in almost every sound mixer existing today. By attenuating the exposed frequencies, the feedback can be eliminated and the overall gain can be raised. When using microphones on a stage for playback on a public address system, the equalization is a very important part of the sound engineer's job. The equalization can also be used for reducing annoying and ill sounding parts of the audio signal.

When feedback occurs, it is highly wanted to remove this feedback without linearly distort the frequency band too much. This means that highly accurate sharp filters should be used to eliminate feedback and these are called notch-filters. A sound engineer will locate the feedback by using an upward gain and increasing the frequency until feedback occurs. Then the notch will be changed to attenuate the frequency until the feedback is gone. This seeking method is even easier if it is combined with a FFT (Fast Fourier Transform) of the signal. By locating the feedback in the frequency spectrum and automatically placing a notch-filter at the exposed frequency, the feedback is killed.

The use of equalization for reducing feedback is something many people have researched. One of these people is Connor [30], who observed an additional gain margin of 10 to 30 dB achieved when using equalizer to reduce feedback, and he also observed that the gain advantage is highly dependent on room form and reverberation. Later has Heinz [8] described an analog version of a tunable notch filter for reducing feedback, and digitally versions have been described by Rocha and Ferreira [31].

In the software of SquareHead's microphone array, the AudioScope, the author of this thesis has created an equalizer in [25], and it can be used to set as many filters as wanted into the signal chain. To implement an adaptive notch-filter box is not very complicated, but the advantage needs to be investigated before an implementation is performed in software.

Today, many commercial products uses the notch filter method to reduce feedback and one of them is the DBX AFS224[2]. This is a multi-notch rack unit that uses 24 sharp notches to kill feedback. By placing the unit in the signal chain, feedback that occurs will be recognized, and a notch-filter will be placed at the exposed frequency. This will continue while the box still has notch-filters left, and after the 24th notch, the oldest notch will be placed at a new frequency. This product will be used when studying the effect of this technique, and it will be determined whether an implementation in software should be made. Since each new box added in the signal chain creates additional delay, the number of extra boxes should be reduced to the absolute minimum. The latency of the AFS224 will be investigated in

[2]http://www.dbxpro.com/AFS224/AFS224.php

section 4.2.3.

One negative effect from the multi-notch anti-feedback box is that it will linearly distort the frequency spectrum. If the filters removes too much of the signal, unnatural sound can be experienced which can be annoying and create low fidelity audio. The expectation from this test is therefore that this box will not work too well alone, but maybe together with other techniques some notches can be used. However, since a standalone box will have its own A/D and D/A converters and its own buffers and processing time, this method should be implemented in software and be performed on the same buffers as the other signal processing to reduce the latency through the system.

The best solution might be to use the software equalizer to remove the worst feedback frequencies and then combine the output with other anti-feedback techniques to achieve the best gain before feedback.

3.1.3 Adaptive Filters

Adaptive filters are used to take away redundancy in channel coding. In coding techniques, this means to find the part of the signal that repeats itself from last sample, and only code the difference between each repeating sample. This results in fewer bits used to describe the signal, and by performing this on the microphone signal, the feedback frequency will be the same for all samples and can therefore be recognized and eliminated.

This method sounds promising, but if speech signals have much redundancy, the signal can be nonlinearly distorted by the algorithm. This can result in unnatural and low fidelity sound. The decorrelation can be performed with a linear predictor and an adaptive filter, but this is a time consuming implementation and much testing is required. The effect will be tested on known feedback cancellers, and the effectiveness of this method need to be proven before an algorithm is created.

An anti-feedback box which uses a decorrelation and adaptive filter algorithm, is the Wideband FC101. This box is mainly made for speech applications and meetings where feedback often occurs and is known to be an effective feedback killer. It will be the test subject to see whether an adaptive filter algorithm is efficient or not.

3.1.4 Spatial Cancellation

Spatial cancellation means to make a spatial band-stop filter in one unwanted direction with the microphones in the array, and this can help to attenuate audio from sources that introduce feedback. This can be difficult to achieve for all frequencies since the directions might be different for every frequency, but for the most vulnerable frequencies this can be useful. When nullifying in live applications, there can be some computation issues that require a powerful computer or some alternative calculation methods that needs investigation. Spatial cancellation might provide better performance, but how this will function is very difficult to predict.

3.2 Delay

Investigations of the delay time are important for studying the effect this has on the overall conference system. Since the signal processing of a microphone array is buffer-based, there will be some delay in the process that is unwanted in a live situation. The sound travel distance will also add up to the delay, and eventually there will be so much delay that it can be audible and annoying. The delay will also be further increased if a loudspeaker array is used in the signal chain. Since there is already a delay from the public address system to the question-asker in the audience with a close-up microphone, the delay deviation from the normal setting must be kept at the minimum. This can be done by reducing buffer sizes in the processing and excluding other possible delay components.

The delay will eventually decide the maximum size of a room the microphone array can be used for. When the echo becomes too audible, a question-asker will have a difficult time to speak. However, the size limit will be the same problem with a close-up microphone because the largest delay is from the loudspeakers to the speaker. In these cases, questions from the audience will be impossible. A subjective evaluation of the delay in rooms needs to be performed to see whether it is possible to use the microphone array or not.

3.2.1 Haas Effect

The human ear is a complex human organ and can tolerate some delay without being disturbed, and something that is called the Haas effect was discovered in the 1970s. The Haas effect is a psychoacoustic effect related to a group of auditory phenomena known as the precedence effect or law of the first wave front. These effects, in conjunction with sensory reactions to other physical differences between perceived sounds, are responsible for the ability of listeners with two ears to accurately localize sounds coming from around them. The effect is first described by Helmut Haas [14].

When two identical sounds are played with different arrival time and sound level, two ears will experience the sound coming from the first arrival source as long as the sound level is not too low compared to the later arriving source. This experience will occur as long as the delayed signal is not more than 30-40 ms compared to the first. After this delay time, the second source will be experienced as an echo dependent on sound level.

When a person is talking in a room, this effect makes sure you hear where the sound originally come from instead of hearing a strong reflection as a source. In a large room, the echo can be experienced if the room is above the Haas limit, which is approximately 15 meters (44 ms). If the second source is above 10 dB louder than the first source, this effect will no longer occur and the louder source will be experienced as the source. The time limit for the delay is related to the level of the amplified loudspeaker.

3.3 Loudness Levels

The signal picked up from the audience must be loud enough to be played on the public address system so everyone can hear it clearly. A conventional close-up microphone will have an advantage here and have a louder sound level. However, the gain that is required in a conference hall need not be out of reach for an array microphone, but this will have a correlation with the gain before feedback. The loudness level achieved from the array will be the final limiting factor, and if it is not loud enough, it will be impossible to use the microphone array.

Tests between a conventional close-up microphone and a microphone array will be created to discover what gain is required in a conference hall. By performing these investigations, different gain requirements can be found for the microphone array.

3.4 Flat Frequency Spectrum

A flat frequency response pickup from the microphone array is required to avoid linear distortion of the signal and to keep the natural timbre of the speaker. The response in the steering direction should have an ideal flat frequency response, but since aperture is affected by steering direction, it will be nearly impossible to have a complete measurement for this because the frequency response is complex to comprehend. A sound source needs to be hit perfectly by the beam to have the best frequency response, and a small off-hit can create differences in the frequency spectrum. One goal is to have the frequency response flat for a zero degrees steering direction. Measurement of the frequency band will point out possible troubles with either the microphone response, the microphone placement, the microphone baffle, or the signal processing response, and eventually design changes or filters can be added to improve the flatness of the response. If there are large linear distortions in the frequency response, masking of speech information can occur which makes the voice harder to understand.

3.5 Optimal Array and Loudspeaker Placement

To improve the overall sound quality and the gain of the system, both the microphone array placement and the positioning of the loudspeakers are critical. Since it is hard to simulate exactly what happens in one specific room, this has to be measured with tests in different rooms.

One possible solution for reducing the overall sound level and increasing the direct sound to the audience, is to use a loudspeaker array in cooperation with the microphone array. Loudspeaker arrays can be used to steer the emitted sound towards the audience and not directly at the microphone array. This can help to increase the gain before feedback level and can be important for optimizing the conference hall for microphone array use [32]. However, the loudspeaker arrays often have a delay, and this disadvantage needs to be weighted versus the advantage the loudspeaker array provides. The loudspeaker array will also give great speech intelligibility at the audience that will help to understand what the speaker is talking about [33].

By performing STI (Speech Transmission Index [34]) measurements, the speech intelligibility of speech captured with the microphone array can be established. This can find information regarding the distance a microphone array can be hung and the degree of speech distortion through the array system.

The human head is not a fully omnidirectional source either, so to pick up the higher frequencies of the speaker, the microphone array has to capture the sound from the front of the talking person. This can be ensured by placing the array in front of the audience steering backwards, as shown in Fig. 3.3. The human mouth's directivity is described in [35].

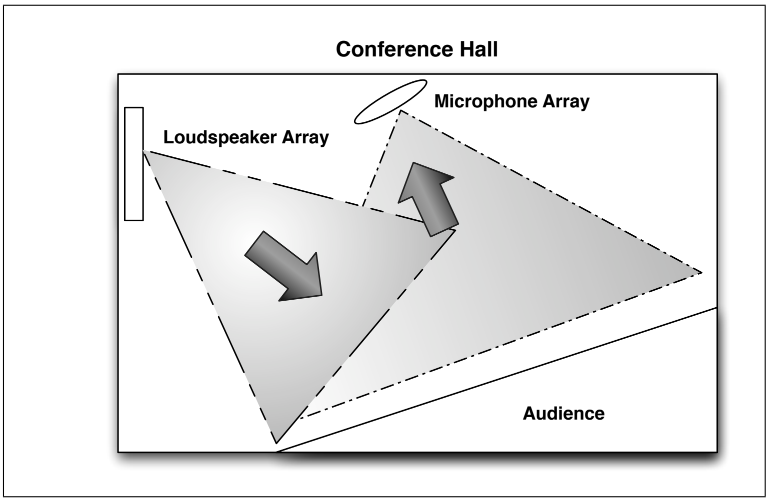


Figure 3.3: Optimal placement in a conference hall

In different conference halls, noise sources such as projectors, ventilation fans, or technical equipment are often present. The array needs to be placed away from these sources so they are not captured. It is also important that entrance doors which are subject to much noise are avoided.

3.6 Automatic Tracking

Automatic tracking is a whole different method for controlling the beam compared with a touchscreen. The touchscreen can be good if there is an operator, but a fully automatic system is a better solution if it is stable. However, there are many pitfalls and these needs to be avoided. Many people have researched microphone arrays with automatic steering and there are varying results [2, 3, 15]. The largest trouble is that the tracking direction will vary very much when there are small noise sources in the tracking room, and these can attract the beam and the wanted sound source will be lost.

The tracking in the AudioScope software will be implemented similar to an ultrasound machine by beamforming and scanning over a selected area. Then one or several frequencies will be analyzed through a DFT (Discrete Fourier Transform), and the sound level from each beam direction will be calculated. The sound level will then be plotted, and the steering will be pointed at the sound level peaks in the plot. The implementation requires much testing, and investigations regarding tracking frequencies, the robustness of keeping a source, the latency of picking out sources, and the attenuation of noise sources, needs to be done.

If the tracking is not functioning well, the tracking can be annoying instead of helpful, and it is important that the tracking will function well in a large noisy auditorium and not only in a small auditorium with little noise. Investigations and algorithms will be tested, to find the most robust tracking method for the described application.

Chapter 4

Measurements and Testing

4.1 Feedback

This chapter describes investigations regarding gain before feedback, and the best performing methods have been established and implemented in software.

4.1.1 Frequency-Shifting

A frequency-shifter can be implemented by splitting the microphone signal into two paths. These paths represent the complex and real part of the signal, and the complex part can be multiplied by a sine and the real part by a cosine in the time domain. In Eq. 4.1 and 4.2, the modulation of the two channels can be seen to separate the real and the complex part from each other.

$$y_1(n) = y(n) \times cos(\frac{2\pi n \cdot Fs}{4 \cdot Fs}) \tag{4.1}$$

$$y_2(n) = y(n) \times sin(\frac{2\pi n \cdot Fs}{4 \cdot Fs}) \tag{4.2}$$

After these processes, the two signals can be filtered by a 63rd order FIR-lowpass filter in Eq. 4.3 and 4.4. This filter will create a 32 sample delay which is about 1 ms in time delay. This filter increases the processing delay of the complete system and is necessary to avoid aliasing.

$$y_1(n) = \sum_{i=0}^{63} y_1(n-i) * firB(i) \tag{4.3}$$

$$y_2(n) = \sum_{i=0}^{63} y_2(n-i) * firB(i) \tag{4.4}$$

After the filtering, the signals will be modulated with the same frequency as in Fig. 4.1 and 4.2, but with a small frequency deviation typically in the range of -5 to 5 Hz, which is the frequency-shift. The math behind these operations can be seen in Eq. 4.5 and 4.6.

$$y_1(n) = y(n) \times cos(\frac{2\pi n \cdot (Fs + shift \times 4)}{4 \cdot Fs}) \tag{4.5}$$

$$y_2(n) = y(n) \times sin(\frac{2\pi n \cdot (Fs + shift \times 4)}{4 \cdot Fs}) \tag{4.6}$$

The two signals can then be summed together in the time domain in Eq. 4.7, and the summed signal has been shifted in frequency by a chosen number of Hertz.

$$y(n) = y_1(n) + y_2(n) \tag{4.7}$$

Fig. 4.1 shows the implementation that is used in the AudioScope software. This implementation was done in collaboration with Ines Hafinovic and Carl-Inge Colombo Nilsen from SquareHead Technology AS.

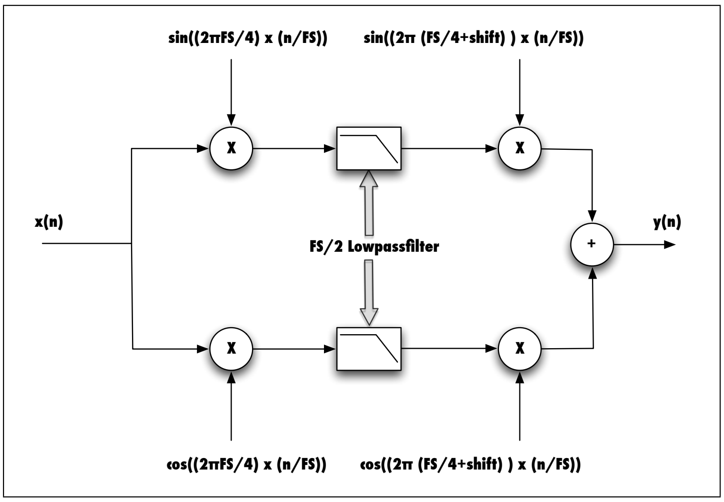


Figure 4.1: Frequency-shifter implementation as performed in this thesis

Test Setup

In this thesis, a setup was created for measuring the gain achieved by adding a frequency-shifter to the signal that was played back on the loudspeaker. The setup was placed in a small auditorium, and this room was 10 meters wide, 12 meters long, and 4 meters high. The audience was tilted so the height at the top of the auditorium was 2.9 meters, and it was 6.2 meters from the front row to the last row. The room was covered with a thin carpet on the floor, and the walls were made of gypsum boards. The ceiling was covered with isolation plates with a 5 cm depth, and there was a movie screen in front of the audience and a small narrow window behind the audience.

The C1 version microphone array had a diameter of 1.55 m, used 285 microphones, and the sampling frequency was 33075 Hz. The array was placed above the second row of the audience with an angle identical to the angle of the audience. This placement ensured that most of the audience was within the microphone arrays camera view and thereby made it easy to select the location where the microphone array should capture sound. Overview pictures from the camera can be studied in Fig. 4.3.

A computer running the application was connected to the microphone array with a cat-5 cable, and the output of the M-Audio Fast Track Pro sound-card was connected to the sound source. The sound source used in this experiment was a Renkus-Heinz Iconyx IC16 which is a digital steerable loudspeaker array and shown in Fig. 4.2(e). The loudspeaker array was programmed with the Beamware[1] software that came with the loudspeaker, and the beamform, which can be studied in Fig. 4.2(f), played only at the audience and not at the ceiling or the floor. The microphone array was placed in the ceiling about 3.4 meters above the audience, and pictures of this setup can be seen in Fig. 4.2.

This setup was created to test gain before feedback, and the level was turned up until feedback occurred (expanding) which is described as gain before howling in Table 4.1. There were also tests of noticeable feedback which is heard as a fading howling in some frequencies of a spoken voice. This is described as gain before noticeable in Table 4.1 and is a subjective evaluation and not as accurate as the howling measurement.

The first test was performed with a bandwidth ranging from 500 to 6500 Hz, and 5 different beams were used to measure the gain before feedback at different locations. The different beam-directions are shown in Fig. 4.3 and were chosen because they give a good estimate of what gain before feedback can be achieved among the audience. When all these positions are studied, an average gain level can be established and a conclusion from the data can be drawn. A test with minor filtering of the frequency spectrum was included which changed the bandwidth from 650 to 6500 Hz with a notch-filter around 800 Hz to reduce the worst feedback.

The first three rows in Table 4.1 were measured without any frequency-shift added to the signal path, and it showed an average of 3.2 dB gain margin from the feedback became noticeable to it started to howl. If a filtering of the frequency spectrum was added, the gain was increased to an average margin of 8.1 dB more than without the filter. If a -5 Hz shift was added, the noticeable gain compared without any shift was averaged to be

[1] http://www.rh.com/support/software-support/iconyx/index.html

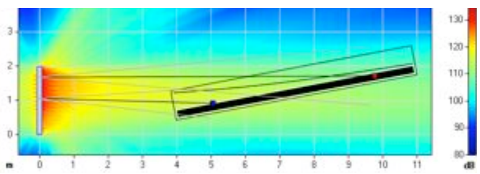
(a) Stage towards the audience

(b) Audience towards stage

(c) Microphone array mounted in the ceiling

(d) Audience and the microphone array

(e) Renkus-Heinz Iconyx loudspeaker array as sound source



(f) Beamform programmed to the Iconyx

Figure 4.2: Frequency-shift measurement setup

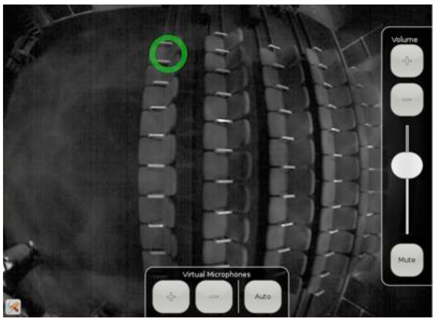
3.9 dB which was also the same as if a +5 Hz shift was added since no level difference in the noticeable gain was shown. From the studies performed by Schroeder [9], the measured gain increase was 6 dB. This was 2.1 dB louder than what was measured in this thesis' setup. However, since the feedback is very dependent on room form and materials, it is difficult to achieve the exact same result as Schroeder did. The conditions are also different since a microphone array and not a single microphone was used.

When the howling feedback was measured, the average gain difference between frequency-shifting and no frequency-shifting was 9.2 dB. This was a measurement which was very close to the theoretical gain of 10 dB that Schroeder [9] found in his simulations. In our setup, this additional gain from 3.9 dB to 9.2 dB was the safety margin before feedback which showed a 5.3 dB gain margin before the noticeable feedback started to howl. If this result was compared with the 3.2 dB margin, which was measured without any frequency-shifter, the margin increase was 2.1 dB, and if the filtering of the band was added, an average margin increase of 2.8 dB can be added. Since the filtering will kill much of the worst feedback frequencies, this extra frequency-shifter after the filter will only give an additional howling margin of 5.5 dB. This will still be a large improvement compared with having no frequency-shifter at all.

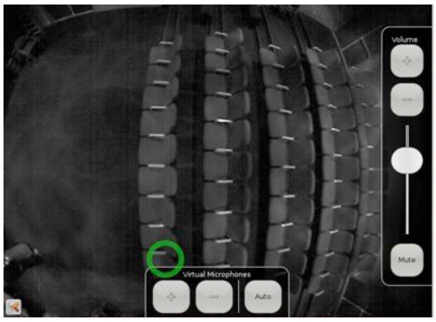
In Table 4.2, the relative gain factor when changing the shifting frequency in steps of 5 Hz can be seen. There was almost 3 dB gain for every 5 Hz that was shifted from the center frequency, but when the shift was above 5 Hz, the frequency-shift effect was audible and will be annoying when talking into the microphone. Since this effect is unacceptable in a conference hall, the main test was performed with a 5 Hz shift only.

The experienced feedback was mostly in the lower frequency range and the -5 Hz shift had therefore a less audible effect when feedback started to develop. The sound effect that the frequency-shift downward produced had a gliding howling like an owl scream, but it was a shorter howling than the high frequency howling produced since the distance in the bandwidth before it became filtered, was shorter. These experienced effects sounded a bit strange, but they can be more acceptable than the system going into feedback and this margin control can be very valuable. However, a musical tone will sound sour through these effects and will limit the use to speech applications.

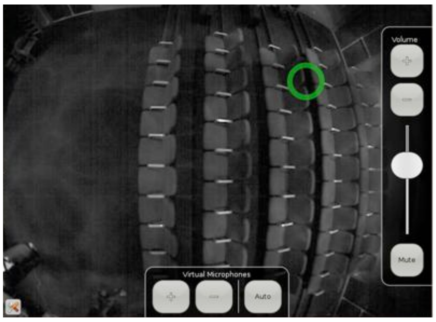
One possible way to solve the long traveling distance in the band that creates the owl scream, can be to have separate frequency-shifts on different sub-arrays. This can make the sound effect filtered faster by the sub-array filters and can be an advantage when suppressing the audible shifting effect. This can be studied in future research.



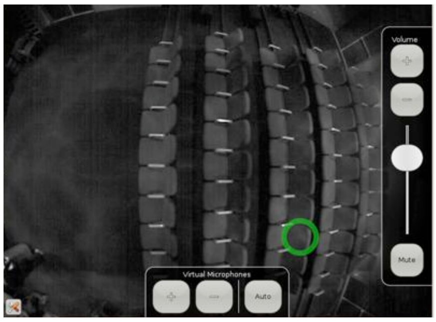
(a) Lower left, 51 degrees to the side



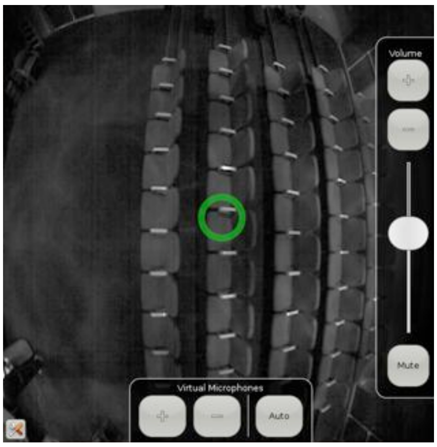
(b) Lower right, 50 degrees to the side



(c) Upper left, 50 degrees to the side



(d) Upper right, 50 degrees to the side



(e) Center

Figure 4.3: Screenshots over different steering positions used in the feedback measurements. The positions are chosen to give a good estimate over the measured gain advantage.

Test type	Middle	Low left	Low right	Up left	Up right
No shift noticeable	−10.5 dB	−11.0 dB	−14.0 dB	−9.0 dB	−8.0 dB
No shift howling	−8.0 dB	−7.5 dB	−12.0 dB	−7.0 dB	−2.0 dB
No shift howling, filter	−1.5 dB	−4.5 dB	−3.0 dB	−1.5 dB	−1.5 dB
-5 Hz shift noticeable	−6.0 dB	−7.5 dB	−9.0 dB	−6.0 dB	−4.5 dB
-5 Hz shift howling	+4.0 dB	+0.0 dB	−3.0 dB	+1.5 dB	+4.0 dB
-5 Hz shift howling, filter	+5.5 dB	+0.0 dB	+1.5 dB	+3.0 dB	+6.0 dB
+5 Hz shift noticeable	−6.0 dB	−7.5 dB	−9.0 dB	−6.0 dB	−4.5 dB
+5 Hz shift howling	+4.0 dB	+0.0 dB	−3.0 dB	+1.5 dB	+5.0 dB
+5 Hz shift howling, filter	+5.5 dB	+0.0 dB	+1.5 dB	+2.5 dB	+6.0 dB
Gain before noticeable	+4.5 dB	+3.5 dB	+5.0 dB	+3.0 dB	+3.5 dB
Gain before howling	+12.0 dB	+7.5 dB	+9.0 dB	+8.5 dB	+9.0 dB
Gain before howling, filter	+7.0 dB	+4.5 dB	+4.5 dB	+4.0 dB	+7.5 dB
Filter gain	+3.2 dB	+1.0 dB	+6.0 dB	+2.7 dB	+1.2 dB
Average, noticeable	+3.9 dB				
Average, howling	+9.2 dB				
Average, howling, filter	+5.5 dB				
Average, filter gain	+2.8 dB				

Table 4.1: Gain before feedback frequency-shift measurements. Noticeable means when noticeable feedback is heard, which means at the edge of feedback. The decibel values are output levels from the software.

Shift in Hz	Relative gain
−5	+0.0 dB
−10	+3.0 dB
−15	+6.0 dB
+15	+4.5 dB
+10	+3.0 dB
+5	+1.5 dB

Table 4.2: Middle steering position with different frequency-shift. Relative gain is measured to -5 Hz frequency-shifter.

4.1.2 Notching and Equalization

In this test, the same setup as the frequency-shifter was used along with the same equipment. The signal chain was also the same except for a DBX AFS224 between the sound-card and the loudspeaker array. In Fig. 4.4, a picture of the AFS224 is shown, and in this setup 21 of the 24 notch filters were used. It was setup by gaining the volume until feedback occurred and the notch-box recognized and killed the feedback.

Figure 4.4: The DBX AFS224 connected to the system loop.

The results from the performed tests can be studied in Table 4.3. The same locations as shown in Fig. 4.3 were used, and the test was performed by gaining the sound level until a clear feedback tone occurred and secondly by noticing feedback on a spoken voice. This was investigated for all 5 directions, and the average gain before feedback was calculated.

When comparing the noticeable level, with and without the AFS224, there was a 6.8 dB higher gain margin with the AFS224. The average gain margin increase before feedback howling, with and without the AFS224, was 11.4 dB. The gain difference between the feedback can be noticed on a spoken voice, to the system goes into feedback had changed from 3.9 dB without the AFS224, to 8.4 dB with the AFS224. This was a large improvement of this system's feedback stability.

Since 24 notches were applied to the frequency band when using the AFS224, there was bound to be large linear distortions in the frequency band. This was something that was audible because the audio sounds thinner and might be too distorted for conference use. However, by using just a few notches at the most exposed frequencies, this can be applied together with other anti-feedback methods to improve a good feedback margin along with acceptable sound quality.

4.1.3 Adaptive Filters

Figure 4.5: The FC101 from Wideband Solutions

Test type	Middle	Low left	Low right	Up left	Up right
Without AFS224 noticeable	−16.5 dB	−18.0 dB	−21.0 dB	−16.5 dB	−16.5 dB
Without AFS224 feedback	−13.5 dB	−15.0 dB	−16.5 dB	−13.5 dB	−10.5 dB
With AFS224 noticeable	−10.5 dB	−12.0 dB	−12.0 dB	−9.0 dB	−10.5 dB
With AFS224 feedback	−0.0 dB	−4.5 dB	−4.5 dB	−1.5 dB	−1.5 dB
Gain before noticeable	+6.0 dB	+6.0 dB	+9.0 dB	+7.0 dB	+6.0 dB
Gain before feedback	+13.5 dB	+10.5 dB	+12.0 dB	+12.0 dB	+9.0 dB
Average noticeable	+6.8 dB				
Average feedback	+11.4 dB				

Table 4.3: Gain before feedback measurements with the multi-notch feedback box connected. The decibel values are software output gain and can be compared with the other values in the table.

The same setup as used in the other feedback tests in this thesis, were used for the adaptive filter test by measuring the gain before feedback with the rack unit connected and without the unit in the chain. Fig. 4.5 shows the FC101 from Wideband Solutions. Since the adaptive filter algorithm was unknown, it was difficult to predict what kind of results that could be expected.

Test type	Middle	Low left	Low right	Up left	Up right
Without FC101 noticeable	−6.0 dB	−6.0 dB	−9.0 dB	−7.5 dB	−7.5 dB
Without FC101 feedback	−0.0 dB	−1.5 dB	−4.5 dB	−1.5 dB	−1.5 dB
With FC101 noticeable	−4.5 dB	−4.5 dB	−9.0 dB	−7.5 dB	−7.5 dB
With FC101 feedback	+1.5 dB	−0.0 dB	−3.0 dB	−0.0 dB	−0.0 dB
Gain before noticeable	+1.5 dB	+1.5 dB	+0.0 dB	+0.0 dB	+0.0 dB
Gain before feedback	+1.5 dB	+1.5 dB	+1.5 dB	+1.5 dB	+1.5 dB
Average noticeable	+0.6 dB				
Average feedback	+1.5 dB				

Table 4.4: Gain before feedback measurements with the FC101 adaptive filter box.

The test results can be seen in Table 4.4, and the adaptive filter algorithm was not performing satisfactorily. There was only a 1.5 dB feedback margin achieved from the box and 0.6 dB noticeable feedback margin. The FC101's algorithm also had a little noticeable nonlinear distortion in the processed signal which made it an even less preferable choice. Since there were no tunable parameteres, improvements could not be made to the test results.

This meant that the conclusion from the adaptive filter test was based on only one algorithm, and it is possible that other adaptive filter algorithms perform better. However, the nonlinear distortion of the frequency band was audible and could be the limiting factor for this method. If there are other anti-feedback methods that does not nonlinearly distort the frequency band as much and perform with the same feedback margin, this algorithm will not be used. Since the results from these simple experiments showed that the feedback margin achieved was very low, it will not be used much effort to implement and test this algorithm, but for future work it might be possible to achieve higher gain margin with a better algorithm.

4.1.4 Spatial Cancellation of Unwanted Direction

Feedback often occured at certain positions in the room, and these locations were characterized as troublesome. One idea was to spatially filter these troublesome areas to increase the gain before feedback level. A spatial filter could be implemented as zeroes in the spatial steering response. This idea could work theoretically, but reflection effects could be as loud as the actual loudspeaker source and thereby erase the advantage completely.

This is an exiting area to study and can be made in future work. The small tests performed by Carl-Inge Colombo Nilsen, Ines Hafinovic and this author have been positive in theory and given a 6 dB gain increase, but the full practical advantage is still unknown.

4.2 Delay

The AudioScope system is described in Fig. 4.6, and all the boxes created some kind of latency through the chain. The delay through each unit was measured and weighted if needed in this thesis, and the minimum possible latency was found.

In this work, the measurement of each chain was performed by sending a test signal from a sound-card. The left output was sent through the box under investigation, and the right output was linked straight to one input of the sound-card. The output from the box was connected to a different sound-card input, and when recording the two channels, the latency was found as the difference in time between the two recorded signals. A MOTU 828 mkII sound-card was used for all the delay measurements which is a professional sound-card with multichannel possibility up to 18 channels simultaneously. Soundtrack Pro was used as the recording software, and the latency was read out in sample difference between the recorded channels. When dividing the sample latency by the sampling frequency, the latency in time can be calculated.

A high attack sound source was needed to get a distinct separation between the sounds, and a drum-like hit was used in these measurement. This sound made it easy to measure the propagation delay through the box under investigation.

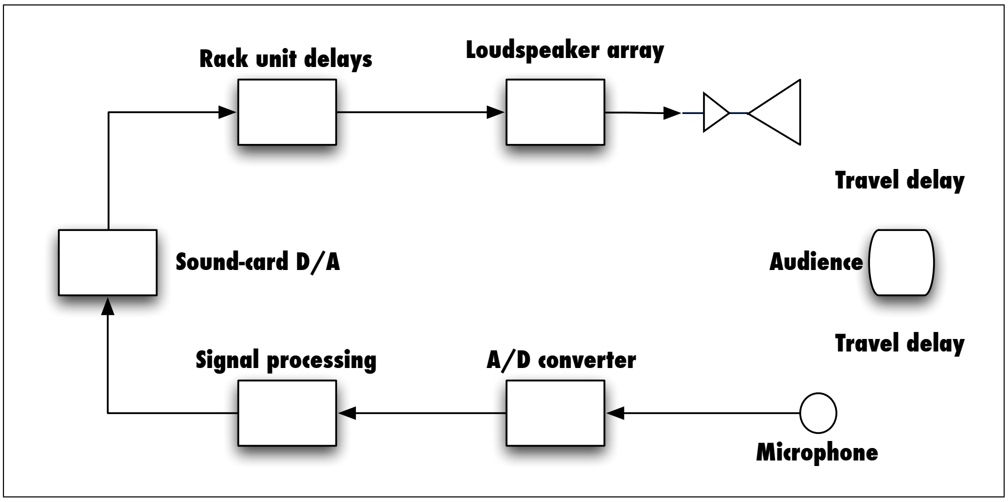


Figure 4.6: System delay chain. Each box represent a delay factor.

4.2.1 Microphone Delay

The microphone used in the investigated AudioScope version was a Pulse DigiSiMic TC100E MEMS (Micro-Electro-Mechanical Systems) microphone with a 4th order sigma-delta modulator. Since the sigma-delta modulator was clocked at 2 MHz frequency, a 10 to 40 sample delay created only 0.005 to 0.02 ms which meant that the delay through the microphone's A/D converter was negligible. In this thesis, measurements of other microphones have been performed as well, but these were of the same type and had therefore the same delay.

4.2.2 Processing Delay

The processing delay was through the software of the AudioScope, and since this software was created especially for the system, the delay through the software could be affected by fast algorithms and smaller buffer sizes. The theoretical minimum latency could be calculated by adding the delay components from the network, processing, and sound-card buffer along with filter delays and signal processing algorithms. The sampling frequency in the processing chain was 33075 Hz.

The network buffers were approximately 2 samples long, and the processing buffers were about 144 samples long. Sub-array filters and frequency-shifting along with equalization, added approximately 70 samples delay, and the sound-card buffer delay was about 96 samples. Summed together, the known theoretical processing delay was 310 samples, which was a delay time of 9.4 ms.

The processing delay started out around 30 ms before the conference application was created and caused echo-effects in the application, but when the code was optimized for speed in this thesis, which included to cut buffer-sizes, shorten filters, and switching sound-card, the total delay was measured to be 10.23 ms. This was performed by cutting the processing buffer size to the minimum which was enough samples over the whole 1.55 meter array. The number of samples were $\frac{1.55 \times 33075}{343} = 150$ which was reduced even further down to 96 by limiting the steering to only 50 degrees and not 90 degrees to the side.

The measured delay was performed with the M-Audio Delta 44 sound-card, and delay measurements were also performed with the M-Audio Fast Track Pro. This sound-card had a measured processing delay of 12.92 ms, which was 2.69 ms longer delay than the Delta 44 sound-card. This difference showed that there can be large deviations in sound-card buffer-sizes which will be reflected in the processing delay, and it is therefore important to chose the correct sound-card for the lowest latency.

There could also be noise troubles with sound-cards related to ground noise or own noise, and this had to be taken into consideration when a sound-card was chosen. Neither of the Delta 44 or the Fast Track Pro showed any special noise troubles in themselves, but the Fast Track Pro experienced some ground noise when used together with one special computer.

4.2.3 Rack Delay

The rack delay was measured through external hardware and should be avoided since most of these boxes had a known algorithm which could be implemented in software. This meant that the extra processing buffers within these rack units were unwanted. However, these boxes could provide great advantages in feedback issues, and extra delay could therefore be accepted. Several boxes were tested to see whether the latency they added could be defended by their advantages.

The tested boxes were:

- Wideband FC101 - Adaptive Filter Feedback Canceller
- DBX AFS224 - Multi Notch Feedback Killer
- DBX 2231 - Analog Equalizer

Wideband FC101 - Adaptive Filter Feedback Canceller

This box provides an adaptive algorithm that suppresses feedback and creates a higher gain margin.

The measured delay through the FC101 was 11.92 ms and was a too large delay for practical solutions in combination with the microphone array. However, this box created some kind of nonlinear distortion on the signal if the volume was close to feedback, and this effect alone excluded it from the chain. Another function this box had, was a noise suppression which reduced the level of noise sent through it, but there were no adjustable parameters to control any of these functions, and the nonlinear distortion these functions created, had a large impact on sound quality that its anti-feedback advantage was lost. The box used an adaptive filter algorithm, and if it is required, it should be implemented into the existing signal processing to reduce the extra latency.

DBX AFS224 - Multi Notch Feedback Killer

The DBX AFS224 is a feedback canceller that detects feedback and place a notch on the exposed frequencies to eliminate the expanding feedback. This box is used in many small feedback prone public address systems and is effective to maintain control of feedback. It can place 24 different notches on each of the two channels, and the notches can have different beam-widths which are adjustable with a knob.

The measured delay through the AFS224 was 0.6 ms which is very low compared with the FC101 feedback canceller. This box can therefore be connected into the signal chain without adding a too large delay, and it can be useful for making a quick setup in a feedback prone environment. Since the filters placed by the AFS224 will linearly distort the signal significantly, the box might not be recommended in a stationary installation, but can be useful when configuring the system.

DBX 2231 - Analog Equalizer

The DBX 2231 is an analog 31 band equalizer which is built up of 1/3 octave band filters where each band has a separate fader. This is a useful tool to find and attenuate feedback prone frequencies quickly in a live sound situation. Since most filters can be implemented in software, this box is not really meant for being in the signal chain, but more as a tool to figure out what filters are required.

The measured latency through the 2231 was 0.1 ms which was a short delay compared with the FC101 and caused by the fact that this was an analog box and had only analog filters and no digital delay. It could probably be used inside the signal chain without too much trouble, but was mostly used to setup a system to find the feedback prone frequencies and add these in software for a final application adjusted for room reverberation.

4.2.4 Travel Delay

The physical distance from the speaker to the microphone array and the distance from the loudspeakers to the audience, are the largest delays in the signal chain. The distance from the loudspeakers to the audience is a natural delay that will be present even if a wireless microphone is used. The distance from the speaker to the microphone array is an additional delay that will affect the overall system, and the microphone array should be as close to the participants as possible for a minimum latency. However, since the microphone array needs distance to be able to steer beams at everyone in the audience, a certain distance from the audience is required. Some travel distance from the audience to the microphone array will therefore be present.

If the microphone array is 5 meters above the question-asker and the loudspeakers are 10 meters from the listening audience, the total travel distance is 15 meters.

$$t = \frac{15m}{343m/s} = 43.7ms \tag{4.8}$$

In Eq. 4.8, the travel time is calculated to be 43.7 ms, and this is longer than the theoretical limit of the Haas effect [14] so an echo might be unavoidable. As the echo will be more noticeable with a longer latency, it will be better to have the smallest delay possible to reduce the echo.

4.2.5 Loudspeaker Array Delay

One solution for stabilizing the system against feedback, is to use loudspeaker line arrays. The advantage of these line arrays is that they do not play audio in the microphone array's direction, in the ceiling, or into floor and walls, which makes the direct sound louder for the audience. However, these loudspeakers have a signal processing chain themselves, and there is a latency from the audio signal is connected to the input until the audio is played out. There are many different loudspeaker arrays on the market, and delay investigations in this thesis have been performed with three of them. According to the producers of the loudspeaker arrays, the delay is dependent on the steering direction, and this will have to be measured for every setup and every time the array is reprogrammed. These delay tests have therefore been performed with the same steering direction and programmed with the same audience area.

There are also some other differences between these loudspeaker line arrays, and these are related to the number of sub-arrays used and the size and alignment of the loudspeaker elements. This will affect the main-lobe width, the side-lobe level, and the grating-lobe level, just as a microphone array.

The first line array studied in this thesis, was the Renkus-Heinz Iconyx 16 which had a measured delay of 6.7 ms. This array was old and a new version of this array will soon be released, but the studied array had only one sub-array because the loudspeaker only used one-sized elements. This created under-sampling for the spatial sources and large grating lobes appeared above 2-3 kHz. These grating lobes created feedback trouble since they hit straight at the microphone array and produced feedback at these frequencies.

The second line array measured was the Tannoy Qflex 32, which was a line array that used two sub-arrays for low and high frequencies to reduce the grating lobe issue. However, the measured delay through the array was 10.7 ms, and this was the longest delay measured through any of the loudspeaker arrays. Even though the sound quality was excellent, the long delay time will be this line array's bane for this application.

The third line array investigated was the EAW DSA250i. This was also a sub-array built up array and was 2-way divided which avoided grating lobes up to 10 kHz. The delay through this array was according to producers, 4.3 ms and the lowest that was found on the market. This meant that this could be the best suited steerable line array for this application.

However, this array has not been tested sufficiently because of failed attempts to borrow a model from the distributors. In future work, this array should be tested in relation with the microphone array to prove its advantages. It is also important that there is a short delay in the loudspeaker array by using passive arrays that can be angled correctly and result in almost zero delay or by using the digitally steerable loudspeaker array with the shortest delay. If the DSA250i is used instead of the Qflex, 6.3 ms is subtracted from the system delay. This will add a few meters to the travel time which will be the largest delay source and therefore the limiting factor.

4.2.6 System Delay

The system delay is the combination of all the mentioned delay sources shown in Fig. 4.6, and the boxes can be added together to find the overall delay. The travel distance creates a delay of 43.7 ms and the processing a 10.2 ms delay. If no rack delay is used and the DSA250i loudspeaker array is used with delay of 4.3 ms, the total delay sums up to 58.2 ms. This is a delay that can be noticed as an echo, but is an improvement of almost 20 ms from the delay that was before the software optimization. There will eventually be a maximum size of halls where the use of microphone arrays will be possible without annoying echoes.

4.3 Frequency Response

When having equipment that records or performs playback of audio, the frequency response of the equipment has a large effect on how the audio will be heard. It is quick and simple to hear the difference between a song played from a PC speaker and a high fidelity system. The PC speaker will often sound thinner than the high fidelity system, and these differences are caused by the lack of bass in the PC speaker's frequency response. It is normal to have equipment that has a non-flat frequency response to make the sound more suited for a special use, and one example of this is headphones with a boosted bass and treble response. The non-flatness of certain equipment is often referred to as coloring of the frequency response. While these examples are highly different from a microphone array, small frequency response peaks can create large linear distortion, and it is therefore important that the frequency response of the microphone array system is as flat as possible for the most natural audio recording.

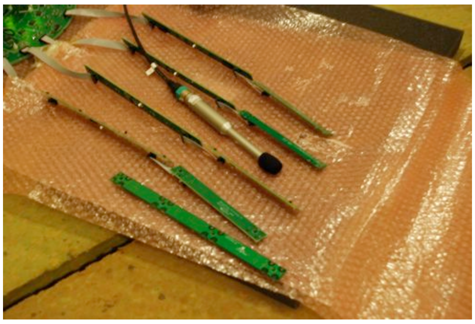
The frequency response of the array microphone is dependent on four parts; the first part is the response of each microphone element, the second is the phase steering and weighting of the elements, the signal processing chain of the audio is the third, and the fourth part is the microphone baffle. When all these four parts are flat, the array microphone frequency response will be flat. The first part is dependent on the microphone used and can therefore vary some for each element of the array, while the next two processes are easy to calculate and can be simulated with imaginary sources and measurement through the steering and filter chain. These two parts can also be used to adjust the frequency response if any large linear distortions are measured. The fourth part, which is the microphone baffle, is the hardest to foresee and can have a large unexpected acoustical effects on the frequency response. It can, however, be somewhat estimated from the Helmholtz resonance.

4.3.1 Microphone Response

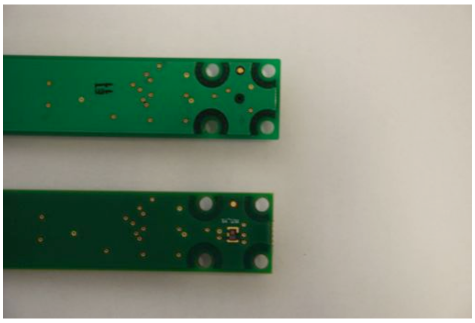
A setup for measuring the frequency response and SNR of seven microphones was assembled. Of these seven microphones, three samples were produced by Analog Devices (ADPM421), which is a new producer, and four were produced by Pulse (DigiSiMic TC100E) which is used in the present design. These two different microphones were chosen because they were MEMS (Micro-Electro-Mechanical Systems) microphones and delivered digital data from the microphone. The main specifications, as described in their datasheets, are listed in Table 4.5. The tests were performed in a semi-anechoic chamber at Department of Physics at the University of Oslo. Fig. 4.7 shows the measurement setup, and the equipment is listed in Table 4.6.

The microphones were mounted on PCBs (Printed Circuit Boards) which were connected to the motherboard and the FPGA that received the microphone data. The FPGA transmitted the audio data to our computers where the data was analyzed and the microphones were listened to. The microphones and the FPGA were supplied from a power supply which was located 1 meter from the microphones. This power supply radiated some 50 Hz noise and was therefore covered with a sound isolation blanket which reduced the noise below audible sound level at the microphone location. The reference microphone was placed in the middle of the test microphones and was connected to one input of the sound-card. The second input of the sound-card

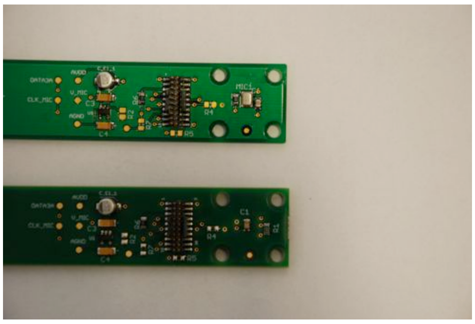
(a) Setup of loudspeaker and microphones



(b) Close up of microphones



(c) Front view: Top is ADPM421, bottom is TC100E



(d) Back view: Top is ADPM421, bottom is TC100E

Figure 4.7: Microphone test setup in a semi-anechoic chamber.

Parameter	TC100E	ADPM421
Modulator	4. order delta sigma	4. order delta sigma
SNR/dB	60 dB	60.5 dB
Sensitivity/dBFS	−26 dB	−26 dB
Vcc (V)	1.64 – 2.86	1.8/2.5/3.3
Current Consumption (μA)	600	650
Clock frequency (MHz)	1 – 3.25	1 – 3.25
Temperature (C)	−30 to +70	−40 to +85
Price/Batch size	EUR 2.35/2k	USD 2.04/1k

Table 4.5: Microphone specifications of the tested microphones. All data is read from the producers datasheets.

was connected from a computer which played audio signal from one test microphone at a time. The loudspeaker was connected to the sound-card output where WinMLS transmitted a 4 second logarithmic chirp while it recorded the two connected inputs. From the plot-view of WinMLS, the frequency responses could be studied and compared, but these measured frequency responses needed to be corrected with the reference microphone, and its response was therefore exported to Matlab for the final frequency response analysis.

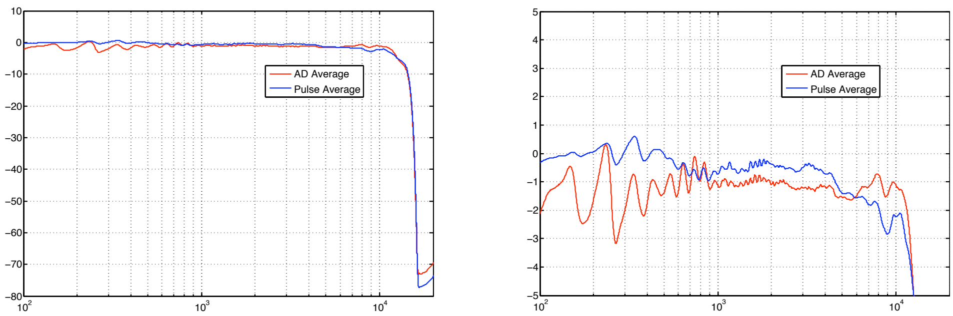
The measurements were performed in three steps where the first step was recorded with one position for microphones and loudspeaker, and the next two steps were with a little different position for both the microphones and the loudspeaker. This small change in position reduced the effect of local variances in the sound pressure, and all three positions were combined in the total average for each microphone. In Fig. 4.8, the average responses of the three Analog Device microphones and the four Pulse microphones are shown.

The frequency response can be studied in Fig. 4.8, and the response was within +/- 2 dB as the producers claimed in their datasheet. The performed measurements showed no significant difference between the two microphones, and the local variation between same type of microphone was larger than the effect of changing microphone type. The ADPM421 microphone seemed to have an oscillation below 1000 Hz which could not be explained by the setup used. This variation was within the specifications of the datasheet, but seemed a little strange since it was the average of all three microphones at three different locations. The most probable cause for the oscillation can be a design flaw in the hardware and for future work this can be studied.

The cut-off frequency of 16 kHz was caused by the sampling frequency at 33075 Hz and the lowpass filter that cut the frequency response. No data was plotted below 100 Hz since this was not relevant in our array application

Type	Name
Reference Microphone	Calibrated Behringer ECM8000
Loudspeaker	ESI nEar 05 Experience
Power Supply	EP-613
Sound-card	Motu 828mk2
Microphone cards	Revision D
FPGA	Firmware 7.108
Software	WinMLS2004 and Matlab
Intern Software	TheApp 7.51
Cables	Assorted
Laptops	1 Windows, 1 Mac

Table 4.6: Equipment used in the frequency response measurements



(a) Microphone responses

(b) Close up of microphone responses

Figure 4.8: Microphone responses, ADPM421 is red and TC100E is blue (x-axis is frequency and y-axis is decibel)

due to insufficient steering at these frequencies.

Microphone	Average Measured Dynamic (SNR)
ADPM421	59.7 dB
TC100E	59.9 dB

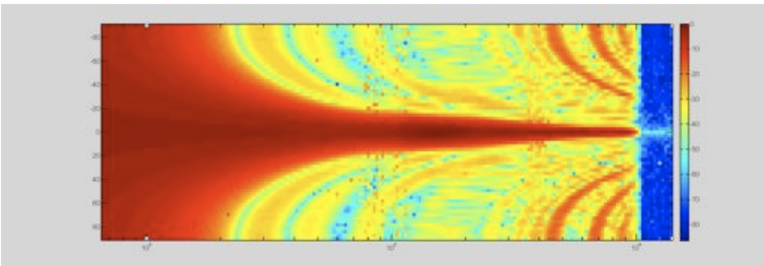
Table 4.7: Measured dynamics of the test microphones

There were also performed measurements of the microphone SNR, and this was tested by having them placed in the semi-anechoic chamber with no sound source. The microphones were recorded and the average peak to peak signal was measured for each microphone, and from this data the dynamic range was calculated. Since the SNR was given in the datasheet, it was expected to lie close to 60 dB. In Table 4.7, the result can be studied and the difference between the measurements and 60 dB was not significant. Since this chamber was only semi-anechoic, there could be some noise in the room that caused the measured SNR to have deviations from the datasheet. The measurements were also performed with the door ajar since cables needed to get out of the room to the recording devices and may have introduced some noise in the measurements.

Since the differences between these microphone types were small, the ADPM421 microphone is cheaper and has a better mechanical solution than the TC100E and may very well be the next microphone used in the array microphone.

4.3.2 Signal Processing Chain

In this thesis, bandwidth measurements were performed through the signal processing chain of the array microphone. This was done by processing all frequencies ranging from 100 Hz to 14 kHz with 200 logarithmic steps and different steering ranging from -90 to 90 degrees in 63 steps. A sound source at 0 degrees and 20 meters in distance was simulated, and there were no filters adjusting the bandwidth other than the sub-array filters that made up the signal processing.



(a) Simulated signal processing chain (x in Hz, y in degrees)

(b) Broadside(0 degrees) frequency response (x in Hz, y in dB)

Figure 4.9: Signal processing response of the microphone array. The results are measured by impressing a sine signal from all directions with stepping frequency.

It can be noticed from Fig. 4.9(a) that at 400 Hz the beam-width was

+/- 20 degrees and at 700 Hz the beam-width was at its normal value of +/- 10 degrees. The low frequency limit was caused by the number of spatial samples and the distance between these. At 4 kHz and above, the main-lobe became more narrow than at lower frequencies, and this was because the sub-array for these frequencies ranged from 3.4 kHz to 10 kHz and thereby created a narrower main lobe for higher frequencies.

Another issue was the grating lobes that appeared from 4 kHz and above, and these were created by the spatial aliasing of the microphone array. In practical situations, some grating lobes can be accepted since there is little high level annoying noise at these higher frequencies, and the upper cut-off frequency can be set at a frequency slightly higher than 4 kHz. The grating lobes can be reduced by adding more microphones closely together in the center of the array.

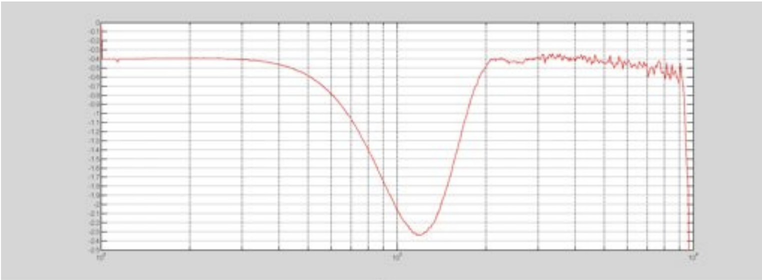
When studying the beam at broadside (0 degrees) in Fig. 4.9(b), the spectrum had an irregularity in the frequency band from 700 - 1100 Hz with a 1 dB attenuation and in the frequencies from 1100 - 2300 Hz with a 2.5 dB gain. The gain will especially make audio sound a bit harsh because of the deviation from a flat response. This was unwanted and needed to be made flat by either an adjustment of sub-array filters, an external filter, or both. The linear distortion in the signal processing chain was created by the cross-over between the sub-arrays.

The sub-array division was done to avoid phase errors in the steering, and this ensured the best possible frequency response from all directions with this weighting. As can be seen from the frequency response measurements performed later in this thesis, the frequency response did not change much between 0 and 45 degrees steering angle.

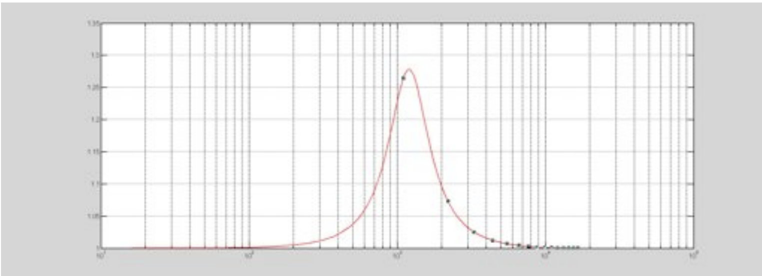
If the sub-arrays were weighted, the beam at broadside will become like Fig. 4.10(a). The weightings were calculated through the simulation software.

- Sub-array 1 ($3400 - 6000$ Hz) = 1
- Sub-array 2 ($1900 - 3400$ Hz) = 0.9225
- Sub-array 3 ($1400 - 1900$ Hz) = 0.5738
- Sub-array 4 ($400 - 1400$ Hz) = 0.9829

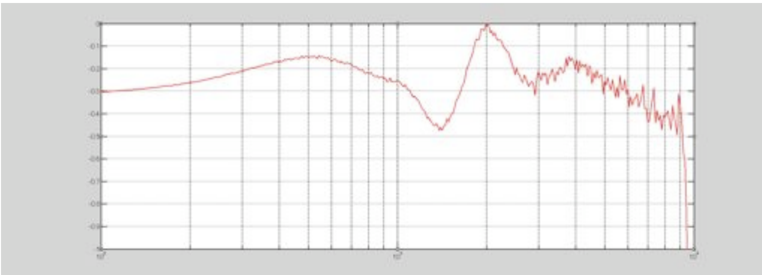
The frequency response peak was now only downward and not both up and down and an external correction filter could be applied to the chain. By studying the dip, a peaking filter with a center at 1200 Hz, a peaking response of around 2 dB, and a bandwidth of $500 - 2000$ Hz, was added. The filter was created using Matlab and added to the signal chain for simulation. The filter was created as a 2nd order infinite impulse response filter, and the response is shown in Fig. 4.10(b). When it was added to the signal



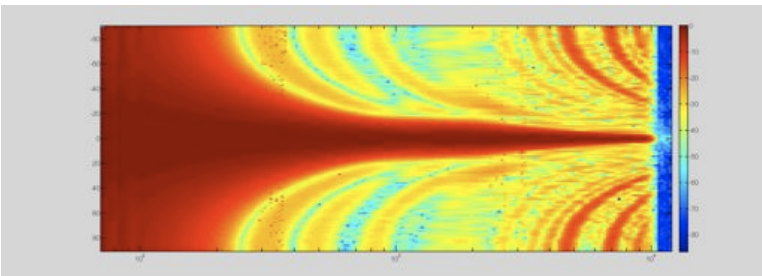
(a) Frequency response after sub-array weighting. (x in Hz, y in dB)



(b) Peaking filter response (x in Hz, y in dB)



(c) Frequency response after sub-array weighting and filter (x in Hz, y in dB)



(d) Frequency response in all directions adjusted (x in Hz, y in degrees)

Figure 4.10: Correction of the frequency response performed in the signal processing chain.

chain, the overall frequency response through the signal processing chain at broadside became like Fig. 4.10(c).

By using sub-array weighting and adjustment filter, the deviation in frequency response was $+/-0.25$ dB. It was difficult to achieve a flatter frequency response because the downward and upward peaks of Fig. 4.10(c) were both affected by filter changes. However, it was a good improvement between the response before and after correction.

When performing a complete run-through of the signal processing chain with an imaginary source at 0 degrees and 20 meters distance, the overall steering simulation became like Fig. 4.10(d). The adjustments that were created have not had any effect on the beam-width of the steering and can be applied in the listening application.

4.3.3 Measured Microphone Array Frequency Response

When both the microphone response was relatively flat and the signal processing chain was flat, the frequency response of the whole microphone array was measured. The frequency response could be affected by the water-protecting material and the microphone's baffle.

The measurements were performed in one small auditorium, one large auditorium, and in a semi-anechoic chamber. These places were chosen because one of them was ideal and the optimal response could be seen, while the others were for studying how an auditorium would affect the frequency response. It was important that the room did not color the frequency response too much and the effect could be compared between the different rooms. The goal was to achieve the most natural frequency response in both the theoretical example and for a special room.

In the small auditorium, the array was mounted in the ceiling pointing with a 20 degrees angle down towards the audience. The ceiling height was 4 meters, and the loudspeaker source was placed in the middle chair about 3.5 meters from the microphone array. The same equipment that was used in the microphone frequency response measurements, was used in this setup. Two measurements at 0 degrees and 45 degrees were performed and the results were averaged over three different recordings. The impulse response was truncated at 7 meters to avoid any large reflections from walls, floors, and ceiling. The truncation also averaged the response.

All the performed frequency response measurements in this thesis were done with WinMLS, which is an acoustic measurement program. This program transmitted a four second logarithmic chirp which was recorded again through the test microphone and the reference microphone. The data was then analyzed and corrected with the reference microphone before the final frequency response was found.

In Fig. 4.11, the measured frequency response of this setup is shown, and it can be seen to be far from flat. The cut-off below 400 Hz was expected

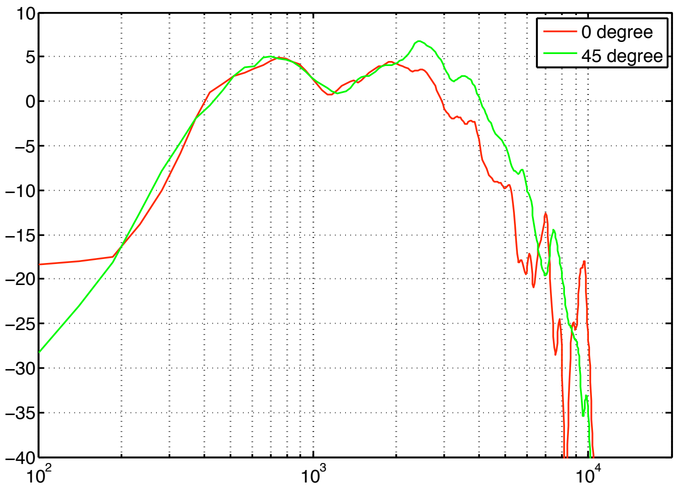


Figure 4.11: Frequency response in a small auditorium with steering angle of 0 and 45 degrees, x-axis is frequency and y-axis is decibel

because of a high-pass filter, but the dip at 1200 Hz of 5 dB was a large irregularity from the other performed measurements. Above 3 kHz, the response fell sharply, and at 4-5 kHz, the attenuation was so large that audio above these frequencies could not be heard properly. These deviations from flat spectrum were large and unexpected, and the reason could be either the room effect or the microphone baffle since the rest of the processing was shown to be flat.

Fig. 4.12 shows the results from the measurements of the large auditorium. Since the distances were large, this impulse response was not truncated and included therefore much reverberation which colored the frequency response. However, the same peak at 2 kHz could be seen in all the three measurements as was also seen in the small auditorium. There was also a peak at 5 kHz which was not seen in the small auditorium's frequency response, and it was difficult to explain why it showed up here. The measurements were made with a different microphone array, but this array was produced the same way and had no large deviations from the other arrays.

Fig. 4.13 was measured in a semi-anechoic chamber and shows the array processing straight ahead at a source that was located 3 meters from the array. As can be studied from this plot, there was a clear peak from 6 kHz and up to the upper cut-off frequency. There was also a clear attenuation in the frequency spectrum from 1000 Hz and below. From 300 Hz and down, a hardware filter caused the attenuation, but the dampening of 300 - 1000 Hz could not be explained by this filter. At 400 Hz, the array had attenuation of about 5 dB compared with the flat area from 1.5 kHz to 6 kHz.

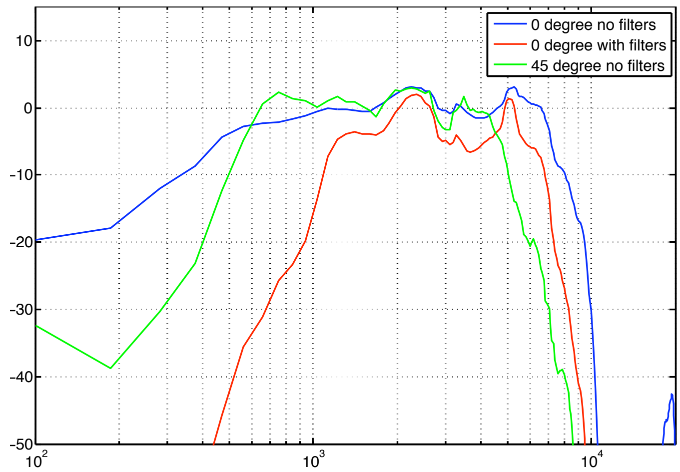


Figure 4.12: Frequency response in a large auditorium with steering angle of 0 and 45 degrees, x-axis is frequency and y-axis is decibel

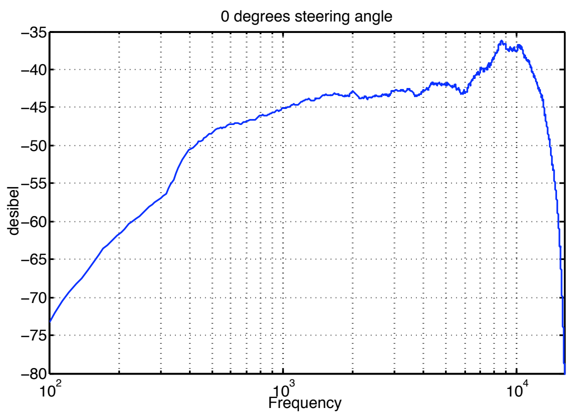


Figure 4.13: Frequency response in a semi-anechoic chamber with steering angle of 0 degrees, x-axis is frequency and y-axis is decibel

4.3.4 Microphone Baffle of the Microphone Array

Since the frequency response from the whole array was somewhat colored and not flat, there could be another reason for coloring the response. Because the microphone's response was flat and the signal processing response was flat, it was only the microphone baffle left. It was possible that the microphone baffle caused resonances that colored the frequency spectrum. This was tested by performing measurements in the same semi-anechoic chamber used in the other measurements.

Helmholtz Resonance

The microphone baffle will create a cavity where a resonance might occur. This resonance is somewhat related to that which Helmholtz found in his resonators and can be read about in [36]. The resonance is dependent on having a small neck and thereby a larger cavity where sound resonance occurs. The Helmholtz resonance formula is

$$f_r = \frac{c}{2\pi}\sqrt{\frac{A}{V \cdot L}}, \tag{4.9}$$

where f_r is the resonance frequency, c is the sound speed in air, A is the neck area, V is the volume of the cavity, and L is the neck length. However, when the neck is small, as can be similar to our case, the L is often changed to $L + \alpha$. According to the microphone producers of Analog Devices, the α can be equal to $\sqrt{\pi}\frac{D}{2}$, where D is the diameter of the neck. If the neck is considered to be circular, the two estimation formulas can be written like

$$f_r = \frac{c \cdot D}{4\sqrt{\pi V L}} \tag{4.10}$$

and

$$f_r = \frac{c \cdot D}{4\sqrt{\pi V(L + \sqrt{\pi}\frac{D}{2})}}. \tag{4.11}$$

These formulas will never give an exact result since the baffles are not shaped exactly spherical or squared, but can provide an estimate of the resonance frequency that can be expected. When the formulas are used in the SquareHead's microphone baffles, the parameters used are; $c = 343$ m/s, $D = 2$ mm, and $L = 1.5$ mm. The volume, V is dependent on shape of the baffle and for each test.

The performed tests were of the old baffle and Pulse microphone with adjustments of the outer hole diameter, new baffle and Pulse microphone with same adjustments of the outer diameter, and a new microphone baffle with the new Analog Devices microphone.

Old Microphone Baffle and Pulse Microphone

The old baffle is shown in Fig. 4.14 and was made up of several layers before reaching the microphone. The outer layer was a carbon layer which had a hole diameter of 4 mm where sound could flow through. The next layer was made from acoustex which is a waterproof and sound transparent material that protects the microphone from moisture. A metal layer was placed underneath the circuit board to make room for the microphone which was surface-mounted on the PCB. This metal layer had a hole diameter of 2 mm where sound waves could travel.

In this design, many small cavities were located around the microphone. These cavities were not ideal for sound pickup because resonances could occur and color the microphone's frequency response. The travel distance from the microphone element to the free air could also create additional resonances similar to that of the human ear.

When estimating the Helmholtz resonance frequency of this microphone baffle, the volume was estimated to be about 273 mm^3. The resulting resonance frequency was 4778 Hz with Eq. 4.10 and 3235 Hz with Eq. 4.11. At least one peak was expected to lie in this area, and there could be additional harmonics to this frequency as well.

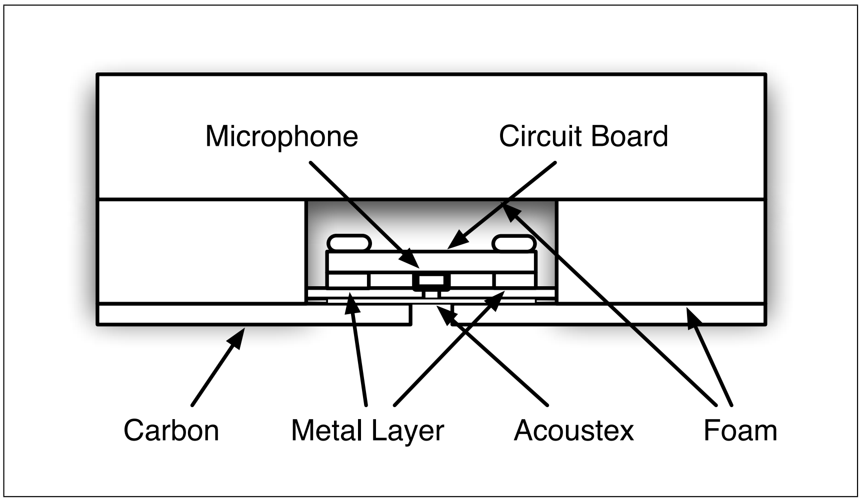


Figure 4.14: Illustration of the old microphone baffle with Pulse microphone.

The frequency response of this old design can be seen in Fig. 4.15, and there were clear resonance peaks in the frequency response. The two worst resonances were at 4.5 kHz with 9 dB peak and at 10 kHz with 11 dB peak. The lowest frequency lie in the area estimated by the Helmholtz equations, while the upper frequency could be the 2nd harmonic. The response showed that there was an increase in gain for the higher frequencies, and this made the frequency response of the microphone array uneven. The linear distortions were very large and could be the cause of the frequency response shown in Fig. 4.13.

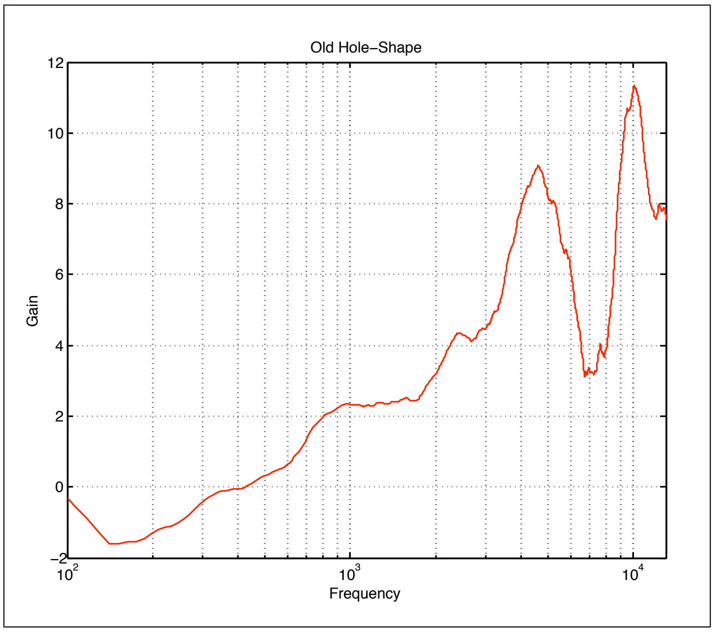


Figure 4.15: Measured frequency response of the old microphone baffle with Pulse Microphone.

New Microphone Baffle and Pulse Microphone

A new baffle was designed by colleagues which was mostly made to cut costs and not created for acoustical advantages. The new baffle is shown in Fig. 4.16 and was basically the same as the old baffle except for the metal layer that was exchanged for an epoxy plastic layer which created a locket around the microphone. This locket had a volume which could create resonances, but there were no large cavities around the microphone which might reduce the experienced resonances or at least place the resonances at higher frequencies.

The volume of the locket could be estimated to be $\pi \cdot 3.6^2$ mm$^2 \cdot 1.5$ mm $= 61$ mm^3. When considering this volume with Eq. 4.10 and Eq. 4.11, the resulting estimated resonances were 10109 Hz and 6844 Hz. This meant that the resonance frequency of this locket would probably be higher in frequencies than the other microphone baffle.

The frequency response measured from this new design can be seen in Fig. 4.17, and there were some resonances in this response. At 6 kHz and 9 kHz there were some peaks which were about 2 dB louder than the surrounding frequencies. In this response, the low frequency response was attenuated more relative to the higher frequencies.

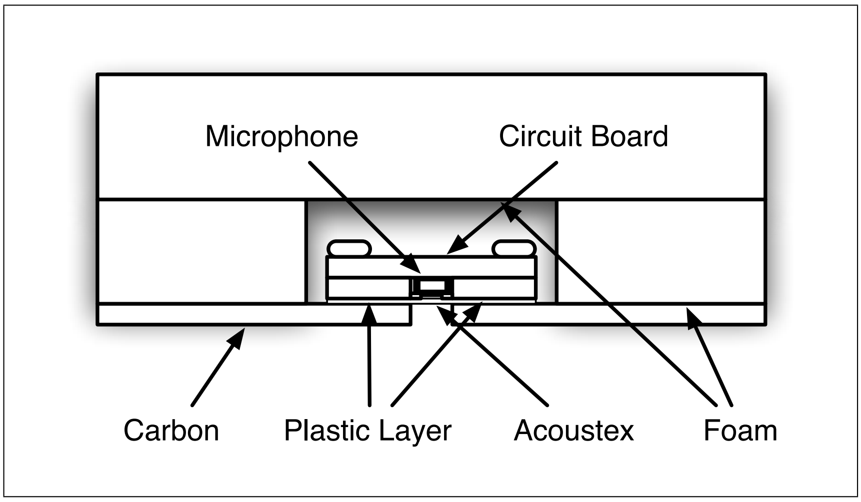


Figure 4.16: Illustration of the new microphone baffle with Pulse Microphone

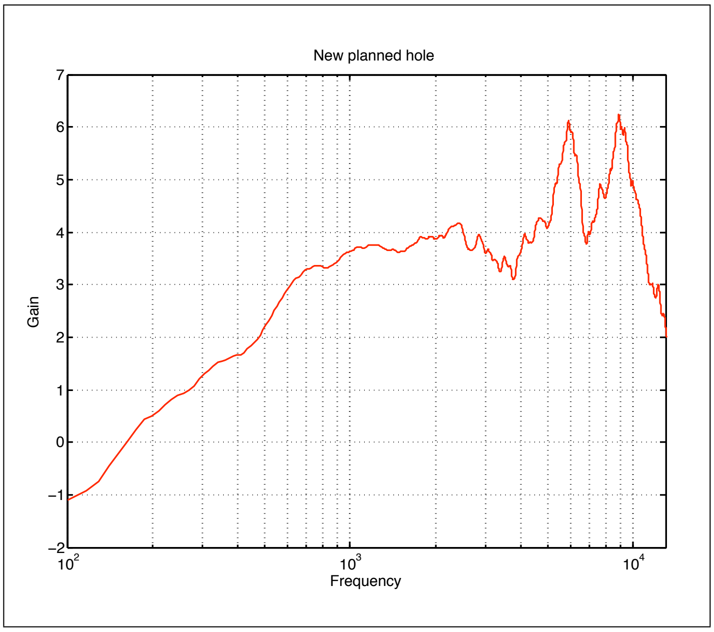


Figure 4.17: Measured frequency response of the new microphone baffle with Pulse microphone.

New Microphone Baffle and Analog Devices Microphone

The new Analog Devices microphone had a different microphone mounting and was placed on the other side of the printed circuit board with a sound transmission hole through the PCB. This made it easier to place the microphone closer to the hole since the PCB could be placed close to the carbon chassis. The baffle can be studied in Fig. 4.18. This new microphone baffle did not create the same resonator as the other baffles did because there was no cavity where the sound could resonance anymore. Since the volume was gone, the Helmholtz equations could no longer be used to predict any resonances. However, if the Helmholtz equations were used, the resonance was estimated to lie above audible level because the cavity was drastically reduced in this new design. The computed resonance frequencies with Eq. 4.10 and Eq. 4.11 and volume equal to 4.71 mm^3, were 36393 Hz and 24639 Hz, and no resonances caused by a cavity should color the frequency response of this new microphone baffle.

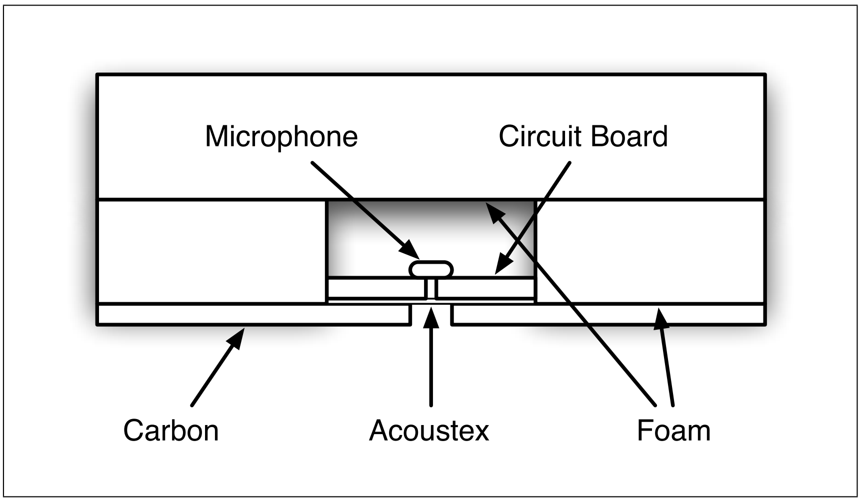


Figure 4.18: Illustration of the new microphone baffle with Analog Devices microphone

When performing frequency response tests with this new hole and microphone, it was surprising to see that the frequency response was almost completely flat as shown in Fig. 4.18. This was much flatter than the other microphone baffles and could be very advantageous acoustically. However, this microphone was not in production yet and could therefore not be tested for a complete array. Since the achieved from this microphone baffle were so good, further testing should be performed to confirm them.

In this new design, there can be an idea to fill the room above the PCB with a sound absorption material which can reduce any resonance reflections from inside the room. This material can also help to provide a pressure on the PCB which can hold it in place so it cannot move, and this is something that also needs further studying when the microphone is in full production.

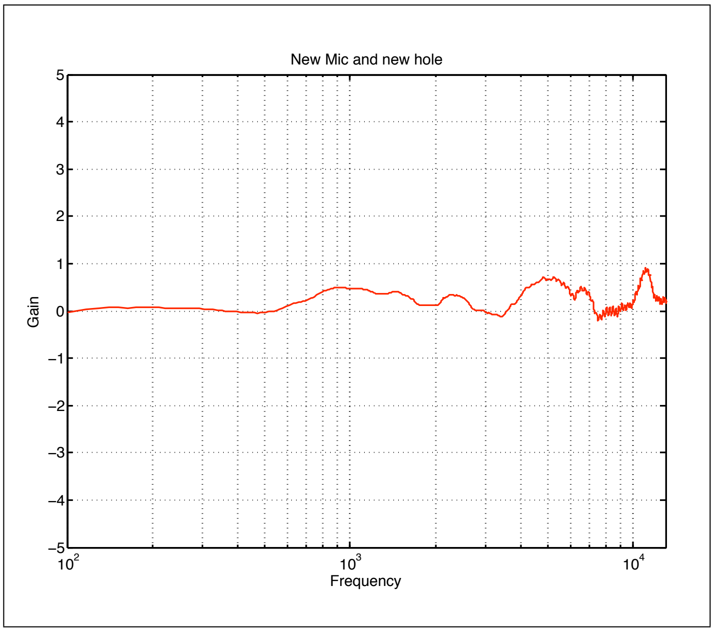


Figure 4.19: Measured frequency response of the new microphone baffle with AD Microphone

4.4 Loudspeaker Measurements

The difference between a conventional loudspeaker and a line array loudspeaker is the directivity. A line array loudspeaker will often have a 30 degrees directivity in the vertical direction, while a conventional loudspeaker will have an omnidirectional directivity at low frequencies and a little sharper directivity at higher frequencies. This difference is something that can be used to focus playback at the audience and not at the microphone array. In a conference hall, the microphone array will be placed in the ceiling above the audience, and if the loudspeaker can be programmed not to play sound at the microphone array, a better gain before feedback margin can be achieved. This is what the next measurements in this thesis are all about.

The test room was the same small auditorium as the feedback measurements were performed in, and the two loudspeakers used, were a Tannoy Qflex32 (Fig. 4.20) line array and a Mackie ART300A (Fig. 4.21) conventional loudspeaker. The Tannoy Qflex32 is a digitally steerable line array that can be programmed to only play sound at the audience. It uses two different element sizes that creates two sub-arrays to avoid the worst grating lobes. This was not the best line array compared in delay time with the

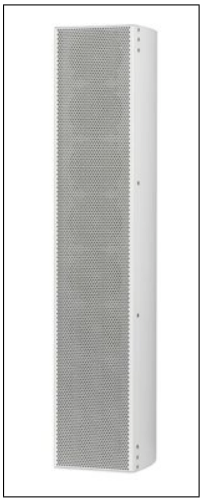


Figure 4.20: Tannoy Qflex 32 loudspeaker array

other line arrays, but since the difference between line arrays and conventional loudspeakers is measured, this array was used in the measurements. The ART 300A is an active regular small PA loudspeaker which is used in many small settings such as indoor installations or events.

Figure 4.21: Mackie ART 300A conventional loudspeaker

Both loudspeakers were placed next to each other pointing at the audience, and they were both calibrated to keep the same sound level with a pinknoise source at the front row of the audience. The sound level from each speaker was measured with a Center 325 Sound Level Meter to be 66 dBA.

A 60 dBA loud pinknoise source (ESI nEar 05 Experience) was then placed among the chair rows in the auditorium. The gain before feedback was then measured with the conventional loudspeaker first and then with the line array loudspeaker. The relative dB difference between the loudspeakers is listed in Table 4.8.

Source	Location of Beam	Sound Level Difference
Pink Noise	Center	4.5 dB
Pink Noise	Side	1.5 dB
Pink Noise	Back	3.0 dB
Music	Center	4.0 dB
Music	Side	1.5 dB
Music	Back	2.5 dB
Speech	Center	4.0 dB
Speech	Side	1.5 dB
Speech	Back	3.0 dB

Table 4.8: Measured relative difference between a conventional loudspeaker and a loudspeaker array. The advantage is for the loudspeaker array in these measurements.

What can be studied from the results was about 4 dB gain level increase with a loudspeaker array in the center of the room, approximately 3 dB at the back of the room, and 1.5 dB when steering to the sides of the room. The level difference between the locations in the auditorium was expected because reflections from walls increased feedback probability, and the closer to the walls the virtual microphone was placed, the smaller level difference was experienced between the two loudspeakers.

Since there were different gain increase throughout the auditorium, it was not possible to read out an exact average gain advantage from the loudspeaker array. There were, however, a gain increase in the whole room that was significant, and the advantage of having loudspeaker arrays cannot be denied.

These decibels that are achieved by using loudspeaker arrays instead of conventional loudspeakers, are of the ”cheap” kind since they do not reduce the sound quality or creates linear or nonlinear distortions in the frequency band. This means that loudspeaker array will help to raise the gain before feedback margin within an acceptable level for conference use.

4.5 Loudness

The loudness test in this thesis was performed in the same small auditorium as the feedback tests were performed in. There was one loudspeaker source (ESI nEar Experience 05) which played a speech sample of 4 seconds. The sample was recorded with the microphone array and a Shure

SM58 close-up microphone before being compared with each other. Since much audio energy was below 100 Hz, which could not be compared because the microphone array could not record these signals, a 4th order butterworth high-pass filter with 150 Hz cut-off frequency was added. This also removed the DC signal from the SM58 recording which gave calculation errors in the sound level.

The array microphone was located in the ceiling steering down towards the loudspeaker which was placed 3.5 meters away. The SM58 microphone was placed 20 cm away from the source when recording, and this is a normal microphone distance when using microphones for speech.

4.5.1 Normal Microphone

The SM58 microphone signal was analyzed with Matlab, and the investigation showed that the SNR between the average of the whole 4 second signal and the silent part was 44.3 dB. The frequency response between the two microphones can be studied in Fig. 4.22.

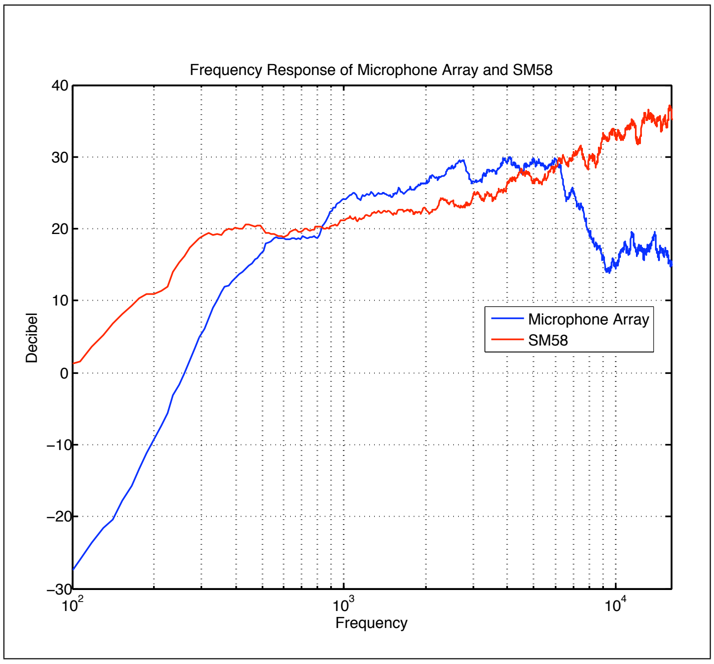


Figure 4.22: Frequency response of the SM58 and the microphone array. Red line is SM58 and blue line is microphone array.

The response was done, since these measurements were performed before WinMLS was bought, by a FFT and adjusted by frequency energy and will be a little inaccurate, but can be compared with each other. The SM58 had a somewhat flatter frequency response and had more energy below 1000 Hz so it sounded rounder. There was also more high frequency energy, but this was because the microphone array's software cut signals above 6.5 kHz.

There was a clear sound quality difference between the two microphones. The microphone array's lack of middle and low frequencies was quite distinct and will be a problem for the experienced sound quality.

4.5.2 Array Microphone

The array microphone had a SNR result of 33.2 dB which was about 11 dB worse than the SM58 close-up microphone. This meant that the microphone array picked up more noise from the surroundings compared with the close-up microphone, but since the distance was 17.5 (12.4 dB) times further from the source this was a quite expected result. However, this meant that the SM58 had 11 dB more gain before feedback at 20 cm from the source than the microphone array at 3.5 m.

If the microphone array should have the same gain before feedback margin as the close-up microphone, these 11 dB needs to be achieved by anti-feedback techniques and signal processing algorithms. This might be impossible, but the closer to the SM58's gain before feedback margin, the better performance will the microphone array experience.

4.5.3 Simple Level Measurements

When the array was tested at a large auditorium, simple decibel measurements were recorded to get the approximately gain level the microphone array system gave. The test was performed with a decibel-meter (Center 325 Sound Level Meter) which measured slow dBA levels. There was one source in the middle of the room, and the decibel-meter was placed at the left front corner approximately 12 meters from the source. The source played straight ahead and not towards the level meter, and this setup resulted in approximately 42 dBA sound level without the microphone array and $52dBA$ sound level with the microphone array. The same test was performed when the source and receiver exchange seats, and this resulted in 39 dBA without microphone array and 47 dBA with microphone array.

This simple test showed that the microphone array gave approximately $8 - 10$ dB sound level increase in the whole room from the speaking source which will increase the possibility of hearing questions from the audience.

4.6 STI Measurements

STI (Speech Transmission Index) [34] measurements were made in this work to find how the array performed in a diffuse noise field. Actually, the measurements were performed with a STI-PA (Speech Transmission Index in

Public Address systems) analyzer because it takes significantly less time than a full STI analysis.

It is important that the speech intelligibility is good when recording the sound source for playback, and this test helps to establish the maximum distance an array can be away from a sound source to be able to understand it in a diffuse noise field. STI measurements are performed by using a test signal at 60 dBA which is a collection of different modulated frequencies. The test signal is a simulation of a speech signal, and by analyzing the recorded signal with the STI meter, a score from 0 to 1 can be established where 1 is the best. The complete STI analysis is based on 14 modulation frequencies with 7 different octave bands, so 98 number of analyses have to be performed. The STI-PA is a little simplified and uses only some of the modulation frequencies for each octave band which reduce the number of analyses to 12. The STI score depends on the content of the recorded signal, and the score can be evaluated according to Table 4.9.

The equipment used were an AL1 Acoustilyzer together with a Calibrated Talkbox source. Both the Acoustilyzer and the Talkbox were produced by NTI[2]. The measurements were performed by having the AudioScope standing vertically and the Talkbox placed at different distances from the AudioScope. The sound output from the AudioScope was connected to the Acoustilyzer for establishing a score. The room was a large football hall, rented from Nydalen Cageball, and several pink-noise sources were placed in the room to create a diffuse sound field of 60 dBA. The Talkbox played the STI signal at exactly 60dBA, and the AudioScope was steered directly at the Talkbox. Pictures from the setup are shown in Fig. 4.23.

STI score	Quality
$1 - 0.75$	Excellent
$0.75 - 0.60$	Good
$0.60 - 0.45$	Fair
$0.45 - 0.30$	Poor
$0.30 - 0$	Unintelligible

Table 4.9: Translation of STI scores to sound quality description.

The goal from these measurements were to establish a maximum distance the AudioScope could function well in. The tests were performed with the whole array, and the results were compared with a single omnidirectional microphone along with the AudioScope simulated in smaller sizes. The results can be read from Table 4.10 where AMD means AudioScope Mounting Dish and the number is the number of microphones used.

The AMD285 and AMD225 were according to the results, quite alike

[2]www.nti-audio.com

(a) AL1 Acoustilyzer

(b) Talkbox

(c) AudioScope dish

(d) Field of measurement

Figure 4.23: STI measurements performed on a football field

Distance (m)	1 microphone	AMD285	AMD225	AMD105
1	-	0.86	0.86	0.81
2	0.40	0.85	0.85	0.77
3	0.33	0.85	0.85	0.71
4	0.29	0.79	0.78	0.64
6	0.20	0.67	0.68	0.52
8	0.15	0.62	0.62	0.47
12	0.10	0.54	0.50	0.39
24	-	0.37	0.35	0.15

Table 4.10: STI-PA measurement results of 8 different distances with 4 different test microphone configurations.

compared to each other. There were only minor differences in the scores, and they were both excellent from 4 meters and below. Up to 8 meters, the score was good and at 12 meters the score was fair. The score at 24 meters was poor, but this distance is unlikely to have in a conference hall. The AMD105 was excellent at 2 meters, was good at 4 meters, fair until 8 meters, and poor at 12 meters. At further distance than 12 meters the STI result showed that it was unintelligible and not usable. Compared with a single microphone, these scores were much higher, and this was expected since the single microphone was poor until 3 meters and unintelligible at longer distance.

4.7 Automatic Tracking

The automatic tracking were implemented in the AudioScope software during this thesis. The tracking were created and adjusted by using subjective try and fail method, and the tracking evolved and changed according to the robustness it showed.

The tracking started out by performing a beamforming grid in the whole video picture and a 4096 point FFT on each beam in the grid. The energy of the sound level was then plotted, and the highest energy was tracked. The result from this test showed that the energy of the speech was not the most optimal way to track because the highest concentration of energy was in the low frequencies below 600 Hz, and the array did not steer too well at these frequencies. In addition to the low frequency steering trouble, there were experienced some troubles because of the fast scanning and the computation time of the FFT. This resulted in reducing the frequencies the sound level were analyzed at. Since the FFT is a computationally heavy transform and unnecessary because only one frequency is analyzed, a DFT for the chosen frequency was used to reduce the number of computations.

The frequency started out at around 1200 Hz, but was found to function better around 1600 Hz depending of gender and timbre of voice, and this became the tracking frequency. Women were often easier to track than men, but there was no real conclusion since this varied for each individual in the test. The chosen frequency represented much of the energy the human voice had in addition to having a narrow directivity from the microphone array, and this combination gave the most robust tracking results. To get an even more robust tracking, one can experiment with tracking several frequencies and weight them accordingly, but has not been done in this thesis.

Since there was one spot in the room which was loudest independent of an existing source or not, a demand that the loudest point was a certain decibel level above the room's average sound level. If the room was stationary, no beam would be focused anywhere if no source was present. This tracking worked well in an ideal room with little noise, but the delay from a person started talking until the person was found by the tracking was too long, and the first two words from a speaker could not be tracked. Another issue was that the tracking found the loudspeaker as a source and put the beam there which resulted in the system going into feedback. There were also troubles where wall reflections from the public address system attracted the beam and not the speaking source in the room.

A tracking area was created to avoid these issues, and the beam was limited to track within this area and thereby excluding the loudspeaker and the wall reflections. This worked well, but not before the maximums on the rim of the tracking area were excluded, which meant that loud sources outside the tracking area could not be tracked. The experience from this version was very good, and in a quiet auditorium, this tracking worked fine except for the delay from the speaker starts talking until the beam was steered at the speaker. When this was tested in a full auditorium with 200 peoples, the result showed that many small noise sources, like a coffee cup being placed on a table, small paper sounds, people stamping their feet, and small talking, were present. These noise sources could easily attract the beam if they were loud enough, and this made the beam go away from the speaking source and thereby making annoying breaks in the speaker's question.

A solution to this trouble was to create two beams. One stayed on the target while the other moved around to the small noise sources that appeared. This helped in keeping the main speaker, but since there now were two virtual microphones in the auditorium, the system became a little unstable for feedback and the level had to be turned down. There were also experienced problems where the question-asker was really low voiced and did not get the beam at all.

The next improvement was to get the tracking faster, and the DFT length was reduced from 4096 to 512 samples. This resulted in a much faster tracking and the delay from the person started talking until the beam

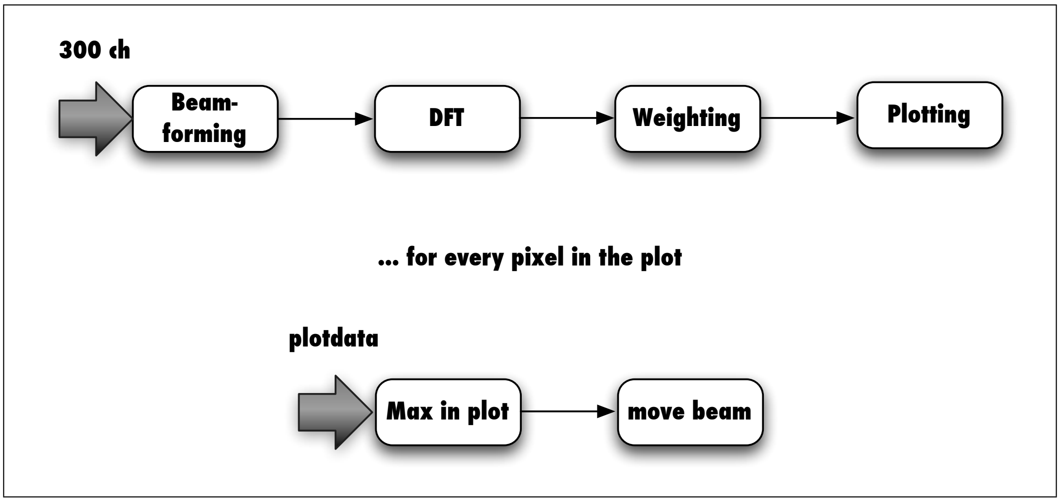


Figure 4.24: The automatic tracking processing chain in this thesis.

was located on that spot, was not a large issue anymore. There were still fluctuations, but the beam would easier come back to the correct person with the faster tracking algorithm.

The tracking is used weekly at StatoilHydro's IB center in Stavanger as this thesis is finished. It is used to find the question-asker and with an override function that the operator use to get full control of the beam. The experience from this auditorium has been very valuable when the tracker has evolved.

4.7.1 Possible Solutions

To find the correct person and to keep the beam, more frequencies need to be tracked, but tests with this have not yet been done. An illustration of the complete tracking processing is shown in Fig. 4.24.

Other investigations have been made to perform the beamforming on a GPU (Graphics Processing Unit) which can be used to do beamforming for tracking on the Graphic card. Carl-Inge Nilsen and Ines Hafinovic have studied this to be faster than doing it on the CPU (Central Processing Unit), and since the GPU uses parallel processing, this has been shown to outperform the CPU. In [37], the method is explained and tested.

Since feedback is dependent on different placement of microphones, the virtual microphone position will decide which frequencies that will howl. This means that a 45 degrees steering to the left will be different from a 30 degrees right steering and can need different parameters. If a matrix is made with different gain and filters to each part of the matrix, this can help to reduce the troubled directions so the overall gain is not dependent on one bad spot. In the middle of the screen, the level can often be louder than near a reflection wall, and this difference can be defined and stored in this matrix. This method can also be used to play the microphone array output onto a multi-loudspeaker setup since different sound can be played at each loudspeaker which optimize the gain before feedback.

This will also ensure that the sources will be weighted with the distance from the microphone array, and a person to the left will be tracked as easily as a person in the center of the audience. An implementation of this matrix will also increase the robustness of the tracker since it can be controlled to not cause feedback. As some directions are more troublesome than others, a person speaking from a troublesome spot might not be gained loud enough for everyone to hear, so improvements of this tracker need to be studied before it can be part of a commercial product. This matrix has been stated, but no measurements have been performed and should be considered in future work.

Chapter 5

Discussion

During this thesis, the performed measurements and results have been analyzed to reach an conclusion. Since many different tests have been done, the most important results have established the working conditions for the use of microphone arrays in conference halls. The conference halls and test rooms have differed in size and shape, and the most natural tested environment was the IB (Visitor) center at StatoilHydro's headquarter at Forus in Stavanger.

5.1 Feedback

The first recognized trouble was the feedback issue. In every microphone live playback system, this issue was present, and even though the studied microphone array had a very sharp capture directivity, it was not protected against feedback. The performed tests in this thesis were mostly based on anti-feedback methods that were implemented in realtime software and afterwards run through a series of different measurements to establish the advantage of each respective method.

The frequency-shifter was one of the first anti-feedback methods that was thoroughly tested in this thesis, and it performed very well to achieve feedback control. This was also what earlier papers had concluded with, and the measured results were close to what the papers about this issue claimed. The feedback margin of the system was increased in this work by 9.2 dB, and was a large improvement and important for an automatic system. The gain before feedback effects were noticed on a spoken voice, was 3.9 dB louder than without this thesis' frequency-shifter. The gain increase enabled the microphone array system to increase the overall gain level by almost 4 dB which was required to be able to play the recorded sound signal loud enough for a large crowd.

The test results were achieved with a frequency-shift of $+/-5$ Hz which were, according to research papers, the maximum shift that was possible without audible nonlinear distortion. There was no measured gain difference between shifting the signal upward or downward and could therefore be chosen according to other demands. There were performed tests with larger shift, and even though these gave higher gain margins, the noticeable nonlinear distortion effect was too annoying for any practical use. The 5 Hz shift also gave audible nonlinear distortion when the gain was close to the limits of the feedback margin, and this nonlinear distortion sounded like an owl scream either upward or downward depending on the sign of the shift. Since most of the experienced feedback was in the low frequency region of 1200 Hz and down, the least audible effect was achieved by shifting -5 Hz and thereby shift the owl scream faster out of the processed bandwidth.

Ideas have been discussed at SquareHead to perform the frequency-shift for separate sub-arrays which are thought to be less audible because the different sub-array filters will filter the owl scream faster. This has not been tested throughly yet and will be considered in future work.

The first filter test was performed with a software equalizer that was created for this purpose by this author's former work, and during this thesis, this was tested in relation with the frequency-shifter. By adding one notch-filter to the most exposed frequency at around 700 Hz, an average filter gain of 2.8 dB was achieved. This gave a clear indication that a frequency band filtering created a large feedback margin, and along with a frequency-shifter, this increased the gain of the signal even louder. The second test was performed with the DBX AFS224 which was an automatic multi-notch feedback killer. The results achieved in this thesis were that an average of 6.8 dB feedback margin before feedback was noticed on a spoken voice and an average of 11.4 dB feedback margin was established from this box before howling began.

These filter results showed a large increase in feedback margin, but there were some troubles with the filtering. The filtering removed parts of the signal and created a narrow sounding voice which made the system sound low fidelity. This reduced sound quality could be too bad for public and commercial use and a mild use of filters was needed to preserve the required sound quality.

The adaptive filter created a feedback margin of only 1.5 dB before howling feedback and 0.6 dB before feedback was noticed which was too low to create the margin needed for microphone arrays.

The spatial filtering has only run through theoretical tests by colleagues at SquareHead, which shows that there are a 6 dB advantage in feedback margin, but since this has not been implemented in a realtime system, full practical tests have not yet been performed.

5.2 Delay

The studies of the delay time through the microphone array system were investigated during this thesis by analyzing each individual part separately. There were numerous sources that produced latency in this system because of the high complexity of hardware and software, and the minimum delay was found through measurements. The goal of this study was to decrease the delay time below the Haas effect, otherwise echoes would occur when speaking into the system.

The measurements showed that the delay time from the microphone and its hardware was almost zero. There were some small delay sources in A/D converters, but since these were small compared with any later latency, they were considered to give zero delay. From the audio was captured by the microphones until the computer had processed it, the delay was measured to be, after software optimization, 10.23 ms. This was very close to the theoretical minimum that was calculated from buffer sizes in the software, and improvements below this delay were hardly possible.

It is possible to change some FIR-filters to IIR-filters to achieve a little less latency, about 1-2 ms, and this is something that can be considered for future work.

As the feedback issue required some anti-feedback killers into the chain, these were also measured for latency in this work. The Wideband FC101, which was an adaptive filter box, had a delay of 11.92 ms, and this box did not perform too well on either the feedback or the delay issue and was not recommended to use. The DBX AFS224, which was an automatic notch-filter box, had a measured delay of 0.6 ms and can be considered for use in the signal chain. The last box was the DBX2231, which was a 31-band equalizer that had a delay of 0.1 ms. This was so low that it could be used if needed, but since the equalization could be done in software, this box was not used in any stationary installations.

The largest delay source was the travel time of sound in air and was the only latency that could not be changed. There were two travel delay sources; one from the audience to the microphone array and another from the loudspeaker to the audience. The travel distance from the loudspeaker to the audience was present in any microphone used, but the travel time from the audience to the microphone array was extra latency. 15 meters total travel time provided 43.7 ms delay, and the only possibility to shorten this, was to place the microphone array closer to the audience which reduced the usable camera view of the array.

The last delay source was the loudspeaker array delay that used latency for beamforming. Three loudspeaker arrays were investigated and measured in this thesis, and these were the Renkus-Heinz Iconyx 16, the Tannoy Qflex 32, and the EAW DSA250i. The respective delay times were 6.7 ms for the Iconyx, 10.7 ms for the Qflex, and DSA250i had 4.3 ms. The DSA250i performed therefore best according to delay and should be used to have the shortest latency. Passive loudspeaker arrays could also be used to reduce the latency even more.

If the whole system delay was added, the delay summed up to 58.2 ms. This was, according to the Haas effect, outside the natural delay limit and would create an echo effect. The strength of this echo would, however, be related to the gain and the delay time above the Haas limit of approximately 40 ms. This meant that any unnecessary delay sources should be cut to get closer to this limit. The possibility of using passive loudspeaker arrays might help along with reducing the FIR-filter lengths.

5.3 Frequency Response

Throughly investigations of the frequency response were done to see if sound reproduction experienced trouble from linear distortions in the frequency response. During this thesis, all elements in the sound capture system, from source to output, were measured to find any deviations from flat frequency spectrum.

The microphones, Pulse TC100E and Analog Devices ADPM421, were measured in free field to have an almost flat frequency response and caused therefore no large linear distortions. They were also both measured to have a dynamic range of approximately 60 dB.

The signal processing was also tested and some linear distortions were found in the frequency response due to sub-array divisions. There was a peak of 2.5 dB at 1800 Hz and a dip of 1 dB at 900 Hz. Since this affects the overall output, adjustments of this frequency response were made, and if any other linear distortions were found, the signal processing's frequency response would be adjusted to correct for these as well.

The shape of a room would affect how a frequency response from a whole array would look like, and tests were therefore performed with ideal conditions in a semi-anechoic chamber. These resulted in a frequency response with some deviations from a completely flat spectrum. There was a peak in the frequency response of approximately $5-6$ dB above 6 kHz which could introduce extra noise from these frequencies, and there was also an attenuation below 1000 Hz. This frequency response was not ideal and should be corrected for the best sound reproduction.

Tests of both a small and large auditorium showed that the reverberation affects the frequency response. There were large deviations from a flat frequency spectrum, and it could be difficult to have a flat frequency response in every room. However, a correction of the frequency response for each room could be possible to ensure the best performance.

Since none of the earlier measurements showed any large linear distortions in the frequency response similar to the deviations measured from the whole array, investigations of the microphone's baffle were performed. The Helmholtz resonance equations showed that some resonances could occur in the microphone baffle because of cavity around the microphone. The microphone baffle, that was used in the microphone array, showed a similar frequency response as the whole array response. There were peaks above 4-5 kHz to above 10 kHz along with attenuation below 1000 Hz which could explain the linear distortions of the frequency response. A test was then created to find out if anything could be done to change the microphone baffle so the linear distortions were avoided. Two new microphone baffle tests were performed, and the best result was found with the new microphone from Analog Devices which was constructed based on the Helmholtz equations to not have a cavity around the microphone. This microphone with a new microphone baffle showed an almost totally flat frequency response and looked very promising for a better array performance.

Tests of the new microphone in a whole array were not performed in this thesis because the microphone was still not in production. This is however, recommended for future work.

5.4 Loudspeaker Considerations

The difference between a loudspeaker array and a conventional loudspeaker was measured in this thesis. In the center of an auditorium, the feedback margin difference was 4 dB in the loudspeaker array's advantage, at the side of the auditorium it was 1.5 dB, and it was 3 dB in the back. This showed that there was a gain increase by using loudspeaker arrays compared to conventional loudspeakers. This gain was very good because it did not distort the sound in any way like anti-feedback techniques did.

There exist many different loudspeaker arrays and some of them were steerable, like the one used in the test, but there were also passive arrays that had a straight forward sharp directivity. Both were usable in the conference setup, but passive arrays had a shorter delay and could be used to reduce the system delay. The effect between passive and steerable loudspeaker arrays were subjective tested, and the delay time was the only noticeable difference. Both steerable and passive loudspeaker arrays can be used to achieve better performance versus feedback.

5.5 Loudness, STI, and Placement of Microphone and Loudspeaker Arrays

The loudness difference between the close-up microphone and the microphone array was measured to be 11 dB at 3.5 meters. This was the gain the microphone array had to provide by applying anti-feedback techniques to perform as a close-up microphone.

STI measurements during this thesis showed that the maximum distance for capturing a good speech signal with the microphone array was 8 meters. Above this distance, the recorded signal was difficult to understand and not too good in a conference. Since the most common height in auditoriums were about 5-6 meters, this limit was maintained, but steering to the side and backwards could get above this limit and the sound quality from these positions could be reduced. The position of the microphone array was therefore important to be able to capture sound from every seat in the audience.

Since a person was not an omnidirectional source at high frequencies, there was an advantage if the microphone array could capture the voice from the front of the person speaking. It was therefore best to have the microphone array over the first few rows, angled backwards towards the audience. This ensured capturing from the front along with getting a short distance to most seats in an auditorium. In addition, this placement made sure the direct reflection from the loudspeakers to the microphone array was not located at the front rows, which increased feedback stability for these seats.

The loudspeakers were also, through experience, found to perform best when mounted with a little height pointing down towards the audience. If the loudspeakers did not play straight into the floor or backwall, but only at the audience, this created the optimal acoustic placement since the microphone array and the loudspeakers played in the same direction. (Fig 3.3)

5.6 Automatic Tracking

Automatic tracking was important to be able to use the microphone array system efficiently, and a tracking algorithm was created and optimized during this work. The tracking located a sound source and played this source automatically out on the public address system. The developed tracker proved stability in small spaces with low noise level, but was not good enough for a standalone commercial product. The tracking lacked something in the robustness and reliability before it could be sold in a product. However, when using the tracker in a quiet room, the tracker performed very well and a natural conversation was easily tracked and captured by the microphone array.

The tracker is used today at StatoilHydro in Stavanger with an operator that controls the beam. As the tracker will be improved, the creation of a fully automatic sound reproduction environment will be possible.

A new development has been started to create a grid for different gain and filters at different steering directions to improve the gain before feedback and the trackers robustness. The advantage is that higher gain can be achieved by using several loudspeakers along with numerous sound outputs. This will also provide a better separation of the audience, and the tracker can be optimized for separating noise and actual sources. More tracking frequencies can also improve the robustness of the tracker, but there are still no results from these undertakings and will be considered in future work.

Chapter 6

Conclusion

An investigation of microphone array use in conference halls has been performed and its feasibility has been studied in this thesis. This has been tested and found to perform well with question and answer sessions in large auditoriums. There is however, a question of experienced sound quality since the bandwidth needs to be narrow to avoid feedback.

The largest issues that are present; is the feedback issue, the delay issue, and the frequency response issue. In this thesis, the feedback margin has been improved by using a soft equalization together with a frequency-shifter of -5 Hz. A loudspeaker array has also been used to maximize the stable gain.

The delay issue still creates some echo effects, but by optimizing the software and elimination of unnecessary external hardware, the delay time have been reduced from above 30 ms to 10.23 ms during this thesis. The total latency has also been reduced to 58.2 ms by changing type of loudspeaker used. The largest delay however, which is the travel time, is still present to set a maximum limit for the auditorium size. To get the total delay below the Haas limit of about 40 ms have proven difficult, but to get close to this limit will be possible.

The frequency response of the investigated microphone array has some large linear distortions which is caused by the Helmoltz resonance in the microphone baffles of the array. Investigations of a better acoustical solution has shown that a new microphone baffle with a new type of microphone shows much smaller linear distortions, but this has not been large scale tested yet since the microphone is still not in production. The microphone change has been planned and will be performed as soon as the production of the new microphone has started.

Test results of the comparison between a close-up microphone and the microphone array shows that the microphone array needs to add a SNR of 11 dB to have the same performance at 3.5 meters. During this thesis, this SNR has been improved by adding together the advantage of the loudspeaker array (3 dB), equalization (3 dB), and frequency-shifter (4 dB). This improvement sums up to an increased gain of approximately 10 dB. To get the array to perform well at longer distances, more gain needs to be provided. Nevertheless, the close-up microphone does sound better because of fair low frequency performance from the microphone array.

STI measurements shows that the maximum distance the array produce good results at is up to 8 meters. This means that an array should not be placed higher than about 6 meters. Experience and tests of different placement of the microphone array shows that the microphone array performs well when placed above the 2nd row angled away from the stage so the center of the microphone array's focus is in the middle of the audience. The best performance is achieved with this position on the microphone array together with directive loudspeaker arrays angled with the same angle as the microphone array. This helps to avoid the sound playing directly into the floor or ceiling.

Along with achieving the optimal performance, a tracker is created in this thesis to make the steering of the virtual microphone automatic. This tracker performs very well in low noise environments and tracks a speaking source very fast. It is used weekly in an auditorium in Stavanger with an operator. However, tests in large auditoriums shows that its robustness is not good enough for a standalone commercial product yet.

6.1 Future Work

Since many issues have been investigated, there has throughout the study emerged new problems, and there has not been enough time to improve these along with the already studied issues. Ideas for solutions to the problems arisen are mentioned in this section, and they can be undertaken in the future to improve the microphone array usage in conference halls.

6.1.1 Array Techniques

There are many different array techniques that can be used to achieve a better array performance, and by adding a higher density in the center of the microphone array, grating lobes can be reduced and thereby achieve a smoother and more controlled high frequency response. In the microphone array used in this thesis, the grating lobes appeared at 4-5 kHz and introduced noise above these frequencies.

Since the bass response is fair for this microphone array because of the array configuration, a spherical version, which both create a larger aperture ($2\pi r$ instead of $2r$) and take advantage of superdirective array beamforming, can be designed. A possible solution is also to make a spherical bass frequency addon to the circular array. This will take advantage of superdirective array beamforming for creating an acceptable directivity at lower frequencies.

Otherwise is it also possible to process in more sub-arrays, especially at high frequencies, to make a smoother main-lobe that has the same beamwidth for all frequencies.

6.1.2 Feedback Techniques

Most known anti-feedback techniques have been investigated, but it is still possible to improve many of these techniques. The adaptive filter algorithm has not been throughly tested, and a better version can be developed than the one already tested. However, since the adaptive filter technique removes part of the signal to achieve feedback stability, the reduced sound quality might be too annoying and not acceptable by an audience.

It is also possible to reduce the frequency-shift howl by doing individual frequency-shift on each sub-array. This can reduce the audible effect, but can also create unnatural sounds and needs to be investigated before implemented.

The spatial filtering that Ines Hafinovic, Carl-Inge Colombo Nilsen and this author is researching, can also provide a gain margin by removing the worst sources for feedback in the room. This is researched as this thesis is finished, and results from this method might provide useful gain levels for the whole system.

6.1.3 Frequency Response and Delay Techniques

In the microphone response of the new Analog Device microphone, an oscillation occurs in the frequency response below 1000 Hz. This might be caused by a design flaw in the hardware and should be investigated so the overall frequency response of the whole array is unaffected. The oscillation is not too distinct with one microphone, but can cause issues when appearing in 285 microphones simultaneously. This new Analog Device microphone need a full array investigation to prove its advantage over the former Pulse microphone.

Since the delay issue is important for reducing the echo effect, the latency of the system should be reduced to the minimum needed. There are still some delay sources that can be further cut, and this is to introduce IIR-filters instead of FIR-filters. Especially in the frequency-shifter, a few milliseconds can be spared.

6.1.4 Improvement of Tracking Algorithm

The tracking is important for getting the system fully automatic, and improvements in the tracking should be studied. There are ideas of a grid in the looking direction of the microphone array to get a better feedback stability. This grid should divide the audience into areas where both gain level can be set individually as well as filters and different signal processing. This can kill the worst areas for feedback without reducing the gain level at every position and make the tracker more robust.

Tracking of more than one frequency can also provide a more robust tracker, and by weighting a number of frequencies according to a human voice, wanted sound sources can be found and amplified.

Bibliography

[1] C. P. Boner and C. R. Boner. Minimizing Feedback in Sound Systems and Room-Ring Modes with Passive Networks. *J. Acoust. Soc. Am.*, 37(1):131–135, January 1965.

[2] J. L. Flanagan. Bandwidth Design for Speech-Seeking Microphone Arrays. *IEEE Int. Conf. Acoust., Speech, Sign. Proc.*, 10:732–735, April 1985.

[3] J. L. Flanagan, J. D. Johnston, R. Zahn, and G. W. Elko. Computer-steered microphone arrays for sound transduction in large rooms. *J. Acoust. Soc. Am.*, 78(5):1508–1518, November 1985.

[4] Y. Tamai, S. Kagami, H. Mizoguchi, K. Sakaya, K. Nagashima, and T. Takano. Circular Microphone Array for Meeting System. *Proc. of IEEE. Sensors*, 2:1100–1105, October 2003.

[5] E. Weinstein, K. Steele, A. Agarwal, and J. Glass. LOUD: A 1020-Node Modular Microphone Array and Beamformer. *Int. Cong. Sound Vibrat.*, 14, July 2007.

[6] H. Fletcher. The Theory of the Operation of the Howling Telephone with Experimental Confirmation. *Bell System Technical Journal*, January 1926.

[7] J. Groth. Digital Signal Processing has made Active Feedback Feedback Suppression a Reality. *The Hearing Journal*, 52(5):32–36, May 1999.

[8] H. K. Heinz. Feedback Elimination Simplified. *Audio Engin. Soc. Preprint 1136, 54th Convention*, May 1976.

[9] M. R. Schroder. Improvement of Acoustic Feedback Stability in Public Address Systems. *J. Acoust. Soc. Am.*, 31(6):851–852, June 1959.

[10] F. J. van der Meulen, S. Kamerling, and C. P. Janse. A new way of Suppressing Acoustic Feedback. *Audio Eng. Soc. Preprint 4735, 104th Convention*, May 1998.

[11] T. van Waterschoot, G. Rombouts, and M. Moonen. On the performance of decorrelation by prefiltering for adaptive feedback cancellation in Public Address systems. *IEEE Benelux Sign. Proc. Sym.*, pages 167–170, January 2004.

[12] J. Alisobhani and S. G. Knorr. Improvement of Acoustic-Feedback Stability by Bandwidth Compression. *IEEE Trans. Acoust., Speech, Sign. Proc.*, 28(6):636–644, December 1980.

[13] K. Kobayashi, K. Furuya, and A. Kataoka. An adaptive microphone array for howling cancellation. *Acoust. Sci. Tech.*, 24(1):45–47, January 2003.

[14] H. Haas. The Influence of a Single Echo on the Audibility of Speech. *J. Audio Eng. Soc.*, 20(2):146–159, March 1972.

[15] W. Kellermann. A Self-steering Digital Microphone Array. *IEEE Int. Conf. Acoust., Speech, Sign. Proc.*, 5:3581–3584, April 1991.

[16] D. C. Moore and I. A. McCowan. Microphone Array Speech Recognition: Experiments on Overlapping Speech in Meetings. *IEEE Int. Conf. Acoust., Speech, Sign. Proc.*, 5:497–500, April 2003.

[17] M. Brandstein and D. Ward (Eds.). *Microphone Arrays*. Springer, 2001.

[18] H. Krim and M. Viberg. Two Decades of Array Signal Processing Research. *IEEE Sign. Proc. Mag.*, 13(4):67–94, July 1996.

[19] Y. Yang, C. Sun, and C. Wan. Theoretical and Experimental Studies on Broadband Constant Beamwidth Beamforming for Circular Arrays. *OCEANS*, 3:1647–1653, September 2003.

[20] P. L. Chu. Superdirective Microphone Array for a Set-Top Videoconference System. *App. Sign. Proc. Audio Acoust.*, pages 235–238, October 1997.

[21] W. Täger. Near Field Superdirectivity (NFSD). *IEEE Int. Conf. Acoust., Speech, Sign. Proc.*, 4:2045–2048, May 1998.

[22] I. Hafizovic, M. Kjølerbakken, and V. Jahr. System Configuration for High Quality Audio Capturing in a Large Microphone Array. *Audio Engin. Soc. 31st Int. Conf.*, 31, June 2007.

[23] I. Hafizovic, C. C. Nilsen, M. Kjølerbakken, and V. Jahr. Design and Implementation of an All-Digital Microphone Array System for Real-Time Speech Acquisition. *Unpublished*, 2007.

[24] I. Hafizovic and C. C. Nilsen. Beamspace Generalized Sidelobe Canceller for Microphone Arrays. *IEEE Work. App. Sign. Proc. Audio Acoust.*, October 2009.

[25] A. Holm. Optimization of Sound Quality from a Steerable Microphone Array. Norwegian University of Science and Technology, December 2008.

[26] M. R. Schroder. Improvements of Acoustic-Feedback Stability by Frequency Shifting. *J. Acoust. Soc. Am.*, 36(9):1718–1724, September 1964.

[27] A. J. Prestigiacomo and D. J. Maclean. A Frequency Shifter for Improving Acoustic Feedback Stability. *Audio Engin. Soc. Preprint 192, 13th Convention*, October 1961.

[28] M. D. Burkhard. A Simplified Frequency Shifter for Improving Acoustic Feedback Stability. *J. Audio Eng. Soc.*, 11(3):234–237, July 1963.

[29] M. H. Jones. Frequency Shifter for "Howl" Suppression. *Wireless World*, July 1973.

[30] W. K. Connor. Theoretical and Practical Considerations in the Equalization of Sound Systems. *J. Audio Eng. Soc.*, 15(2):194–198, April 1967.

[31] A. F. Rocha and A. J. S. Ferreira. An Accurate Method of Detection and Cancellation of Multiple Acoustic Feedbacks. *Audio Eng. Soc. Preprint 6335, 118th Convention*, May 2005.

[32] G. W. Elko and M. M. Goodwin. Beam Dithering: Acoustic Feedback Control using a Modulated-Directivity Loudspeaker Array. *IEEE Int. Conf. Acoust., Speech, Sign. Proc.*, 1:173–176, April 1993.

[33] H. Nomura and M. Tohyama. Loudspeaker Arrays for Improving Speech Intelligibility in a Reverberant Space. *J. Audio Eng. Soc.*, 39(5):338–343, May 1991.

[34] H. J. M. Steeneken and T. Houtgast. A Physical Method for Measuring Speech-Transmission Quality. *J. Acoust. Soc. Am.*, 67(1):318–326, January 1980.

[35] F. Bozzoli and A. Farina. Directivity balloons of real and artificial mouth simulators for measurement of the Speech Transmission Index. *Audio Eng. Soc. 115th Int. Conf.*, October 2003.

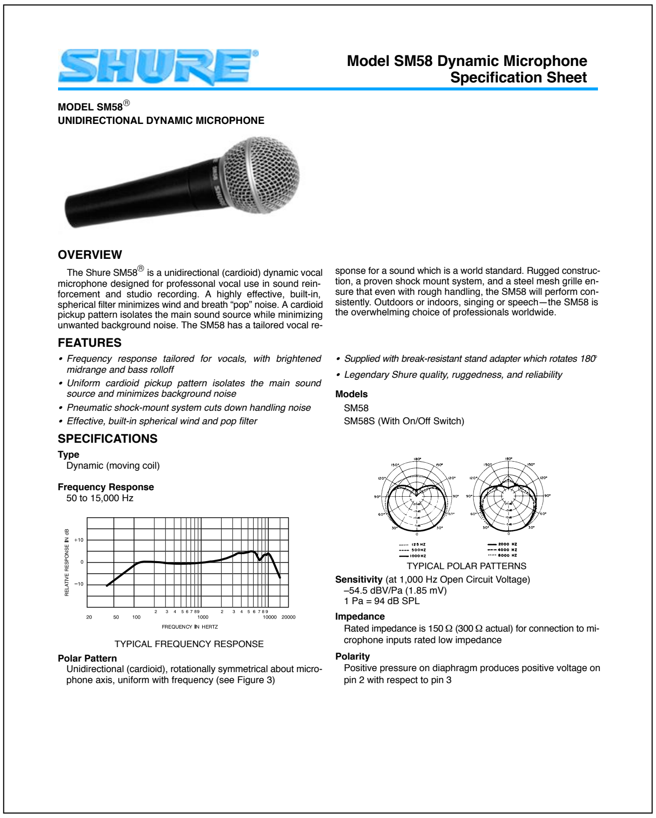
[36] Rayleigh. *Theory of Sound*. New York: Dover Publications, 1877.

[37] C. C. Nilsen and I. Hafizovic. Digital Beamforming Using a GPU. *IEEE Int. Conf. Acoust., Speech, Sign. Proc.*, April 2009.

Appendix A

Datasheets

A.1 Shure SM58



Model SM58 Dynamic Microphone
Specification Sheet

MODEL SM58®
UNIDIRECTIONAL DYNAMIC MICROPHONE

OVERVIEW

The Shure SM58® is a unidirectional (cardioid) dynamic vocal microphone designed for professonal vocal use in sound reinforcement and studio recording. A highly effective, built-in, spherical filter minimizes wind and breath "pop" noise. A cardioid pickup pattern isolates the main sound source while minimizing unwanted background noise. The SM58 has a tailored vocal response for a sound which is a world standard. Rugged construction, a proven shock mount system, and a steel mesh grille ensure that even with rough handling, the SM58 will perform consistently. Outdoors or indoors, singing or speech—the SM58 is the overwhelming choice of professionals worldwide.

FEATURES

- *Frequency response tailored for vocals, with brightened midrange and bass rolloff*
- *Uniform cardioid pickup pattern isolates the main sound source and minimizes background noise*
- *Pneumatic shock-mount system cuts down handling noise*
- *Effective, built-in spherical wind and pop filter*
- *Supplied with break-resistant stand adapter which rotates 180°*
- *Legendary Shure quality, ruggedness, and reliability*

Models
SM58
SM58S (With On/Off Switch)

SPECIFICATIONS

Type
Dynamic (moving coil)

Frequency Response
50 to 15,000 Hz

TYPICAL FREQUENCY RESPONSE

Polar Pattern
Unidirectional (cardioid), rotationally symmetrical about microphone axis, uniform with frequency (see Figure 3)

TYPICAL POLAR PATTERNS

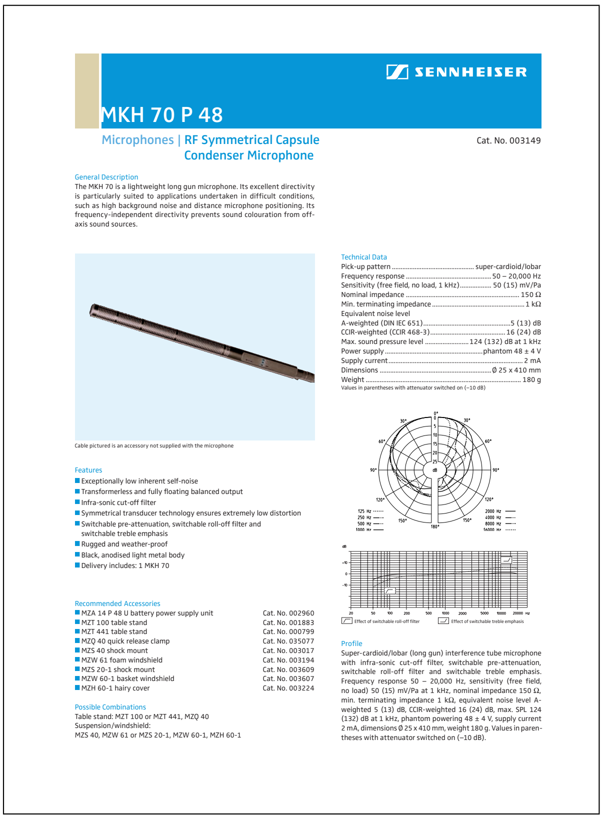
Sensitivity (at 1,000 Hz Open Circuit Voltage)
–54.5 dBV/Pa (1.85 mV)
1 Pa = 94 dB SPL

Impedance
Rated impedance is 150 Ω (300 Ω actual) for connection to microphone inputs rated low impedance

Polarity
Positive pressure on diaphragm produces positive voltage on pin 2 with respect to pin 3

Figure A.1: Datasheet of Shure SM58

A.2 Sennheiser MKH70



SENNHEISER

MKH 70 P 48

Microphones | RF Symmetrical Capsule Condenser Microphone

Cat. No. 003149

General Description

The MKH 70 is a lightweight long gun microphone. Its excellent directivity is particularly suited to applications undertaken in difficult conditions, such as high background noise and distance microphone positioning. Its frequency-independent directivity prevents sound colouration from off-axis sound sources.

Cable pictured is an accessory not supplied with the microphone

Features

- Exceptionally low inherent self-noise
- Transformerless and fully floating balanced output
- Infra-sonic cut-off filter
- Symmetrical transducer technology ensures extremely low distortion
- Switchable pre-attenuation, switchable roll-off filter and switchable treble emphasis
- Rugged and weather-proof
- Black, anodised light metal body
- Delivery includes: 1 MKH 70

Recommended Accessories

MZA 14 P 48 U battery power supply unit	Cat. No. 002960
MZT 100 table stand	Cat. No. 001883
MZT 441 table stand	Cat. No. 000799
MZQ 40 quick release clamp	Cat. No. 035077
MZS 40 shock mount	Cat. No. 003017
MZW 61 foam windshield	Cat. No. 003194
MZS 20-1 shock mount	Cat. No. 003609
MZW 60-1 basket windshield	Cat. No. 003607
MZH 60-1 hairy cover	Cat. No. 003224

Possible Combinations

Table stand: MZT 100 or MZT 441, MZQ 40
Suspension/windshield:
MZS 40, MZW 61 or MZS 20-1, MZW 60-1, MZH 60-1

Technical Data

Pick-up pattern	super-cardioid/lobar
Frequency response	50 – 20,000 Hz
Sensitivity (free field, no load, 1 kHz)	50 (15) mV/Pa
Nominal impedance	150 Ω
Min. terminating impedance	1 kΩ
Equivalent noise level	
A-weighted (DIN IEC 651)	5 (13) dB
CCIR-weighted (CCIR 468-3)	16 (24) dB
Max. sound pressure level	124 (132) dB at 1 kHz
Power supply	phantom 48 ± 4 V
Supply current	2 mA
Dimensions	Ø 25 x 410 mm
Weight	180 g

Values in parentheses with attenuator switched on (–10 dB)

Profile

Super-cardioid/lobar (long gun) interference tube microphone with infra-sonic cut-off filter, switchable pre-attenuation, switchable roll-off filter and switchable treble emphasis. Frequency response 50 – 20,000 Hz, sensitivity (free field, no load) 50 (15) mV/Pa at 1 kHz, nominal impedance 150 Ω, min. terminating impedance 1 kΩ, equivalent noise level A-weighted 5 (13) dB, CCIR-weighted 16 (24) dB, max. SPL 124 (132) dB at 1 kHz, phantom powering 48 ± 4 V, supply current 2 mA, dimensions Ø 25 x 410 mm, weight 180 g. Values in parentheses with attenuator switched on (–10 dB).

Figure A.2: Datasheet of Sennheiser MKH70