Loudspeakers as recording devices in public address systems

Master Thesis Report

Tobias Roest

Technische Universiteit Delft

TU Delft

Loudspeakers as recording devices in public address systems

Master Thesis Report

by

Tobias Roest

to obtain the degree of Master of Science

at the Delft University of Technology,

to be defended publicly on Monday January 24th, 2022 at 13:00.

Student number:	4233050	
Project duration:	May 1, 2021 – January 24, 2022	
Thesis committee:	dr. ir. R. C. Hendriks,	TU Delft, chair
	dr. M. Mastrangeli	TU Delft
	dr. J. Martinez	TU Delft, daily supervisor
	ir. H. Oosterom	Bosch Security Systems B.V.

An electronic version of this thesis is available at `http://repository.tudelft.nl/`.

TU Delft

I wonder if anyone is listening

Abstract

Public address (PA) systems are an integral part of every building. Their function as a safety and security system could benefit from adding recording capabilities. With such capabilities, people could be located during evacuation or announcements could be adjusted to room acoustics for better intelligibility. However, adding microphones to a PA system is too costly and inefficient to be feasible. This thesis evaluates the feasibility of using the loudspeakers already in place as recording devices to provide information to the system. To this end, a system using a single loudspeaker as both a playback and recording device is analysed, modelled and simulated. The results show that using a current measuring set-up with an analogue-to-digital converter capable of detecting a range of roughly 120 dB, a speech signal up to three metres in a cone of 120° from a loudspeaker can be successfully estimated in an office room with an announcement playing and background noise present. As the estimated signal is unknown to the system, the solution generalises to other signal types as well. A system with a single loudspeaker can be utilised for the use cases presented, and is therefore proven feasible. To increase the practicality of the system, it is recommended to continue the research in two main areas. The first area considers improving the quality of the recording and extending the range from which the system can produce accurate recordings. The second area considers of evaluating the practical implementation of the system, by extending the single loudspeaker case to a multiple loudspeaker case and generalising the system to be readily implementable in a large variety of PA systems.

Preface

Since as long as I can remember, I have been enjoying music coming from loudspeakers. When I got my first keyboard at 4 years old, the sounds I could make with it blew my young mind. As an enabler for my life-long passion of music, loudspeakers played an integral role in my life growing up, albeit mostly passively in the background. They allowed me to explore music from across the world and to make music by myself. When I got a bit older, my interest for the production of music grew as well. Loudspeakers and headphones were a part of my experience with sound I became increasingly aware of. At the age of 15, loudspeakers were complemented by my first microphone, confronting me with the reality of my adolescent voice.

During my studies in Electrical Engineering at the TU Delft, my passion for music and sounds grew even further through my increasing collection of synthesizers and understanding of electrical engineering. More abstract sounds and the processing of these became a large part of my life, and I managed to integrate my hobbies with my education as an engineer in a profound bond.

Now, at the end of my studies, I present this thesis report. It contains the results of my research on a subject I feel a deep personal connection with. With a broad scope, ranging from acoustics to electronics design to signal modelling and audio processing, I think this thesis is not just a scientific work, but also a representation of my own affinity with the subject. I have spent the last 9 months researching a creative use of the devices that I have become so accustomed with over the course of my life, and I am proud to present this work as my own.

Over the course over this project I have enjoyed working with many great people, and I'd like to thank these people here.

dr. ir. R.C. Hendriks - Thank you for giving me the opportunity to work on a project that I feel so connected with, and for allowing me the freedom to pursue the research in the directions I thought were most interesting.

dr. J. Martinez - Thank you for going above and beyond to accommodate my research. You were never to busy for an update, and you helped me a lot navigating the bureaucracy the pandemic brought to the TU Delft.

Demi Driessen - Thank you for being my awesome girlfriend and making me smile, even in the most stressful moments.

My parents Ada Brokers and Nico Roest - Thank you for kindling my love of music, and always supporting me throughout my studies.

My brother Max Roest - Thank you for always being ready to help me, and being the best sounding board one could hope for.

ir. H. Oosterom - Thank you for providing me with valuable insights on PA systems and giving me access to resources from Bosch Security Systems B.V.

ir. C. Tchang - Thank you for allowing me to perform measurements in your acoustic lab at Bosch, and helping me with the set-up and devices.

dr. M. Mastrangeli - Thank you for always being enthusiastic. Though we didn't speak a lot during my project, every time I visited you you were very eager to hear about the progress, which was very inspiring.

Ids, Ruben, Michael and the other people at CAS - Thank you for providing the much needed social and academic interactions at the office and outside.

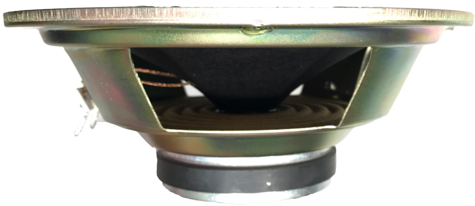
Tobias Roest
Delft, January 2022

Contents

1 Introduction

Across the world, sound reproduction systems are almost omnipresent. From people talking on their phone to musicians being amplified in stadiums filled with thousands of people, sound is being recorded and played back through electronics. In this process, two types of transducers are essential. Microphones convert acoustic signals into electronic signals to be transmitted or stored. The other, loudspeakers, convert an electric signal into an acoustic signal. Over the past few decades, both of these devices have been optimised for their intended direction of transduction, and many different types of loudspeakers and microphones were developed for specific use cases [1–4].

As for loudspeakers, one type still remains the most popular since its first introduction in 1925: the moving-coil electrodynamic loudspeaker. Even the electrical symbol for loudspeakers is derived from the general shape of these types of loudspeakers [5]. An example of a moving-coil electrodynamic loudspeaker is shown in Figure 1.1.



(a) Side view

(b) Front view

Figure 1.1: Example of a moving-coil electrodynamic loudspeaker.

In the microphone realm, several different types are commonly utilised depending on the use case. One of these types is the dynamic microphone, also known as the electrodynamic microphone or the moving-coil microphone. While it is not very common in most applications [4], it is the most popular type of microphone for on-stage singing. As such, most people would recognise the example shown in Figure 1.2.

From the nomenclature of these devices alone, it can be devised that the operational mechanics inside are related to each other. If the loudspeaker and microphone mechanisms are similar enough, perhaps loudspeakers could also be used as microphones. Cases exist where loudspeakers are indeed used as microphones [6, 7], but transparent scientific research is still lacking. This thesis attempts to start closing the gap, by evaluating the usage of loudspeakers for audio recording purposes.

In Chapter 2 possible use cases for recording audio with loudspeakers will be discussed. A reflection on these use cases will allow for proper scoping of the research. It will then continue with the the research objective, as well as the underlying research questions from these cases. After this, the organisation of this thesis will be explained using the research questions.

Figure 1.2: The SM58® dynamic microphone.

2

Research objective

For different types of sound reproduction systems, different characteristics are required from the devices in use. To facilitate this, loudspeakers come in many different shapes and sizes. Even within one enclosure, different sizes of loudspeakers can often be found. A proper scoping and an evaluation of different use cases is therefore necessary. In what case could it be beneficial to use a loudspeaker for recording? This chapter describes such a use case, from which the research objective and research questions will be derived. It concludes with discussing the structure of this thesis.

2.1. Why not use a microphone?

Within the project context, one of the first suggestions to come to mind is to simply use a microphone. Nowadays, small and cheap microphones are readily available, and they have a sufficient quality for most use cases. This is also why they can be found in many systems used in day-to-day life. Smartphones and tablets, laptop computers, and smart speakers all have small microphones inside. Wireless earbuds and hearing aids even perform beamforming techniques with multiple small microphones inside the devices to improve sound quality and speech intelligibility [8, 9]. Almost every system that includes a microphone can record a sufficient quality audio signal for its use. On top of that, for a lot of systems microphones could be added at relatively low cost. For this thesis to have the most impact, it focuses on applications with no microphones present, and where adding microphones would be costly.

A family of systems that fits these needs is the public address (PA) system. These systems are designed for electronic sound amplification and distribution. Within this family, the systems focused on distribution are of particular interest for this project, as the speakers and the microphones are isolated from each other. These PA systems are the ones found in, e.g., office buildings, hotels, or train stations, where the focus is on announcements and voice alarming. An example of a PA loudspeaker integrated in a ceiling is shown in Figure 2.1. These systems are partly embedded into buildings during construction. Changing the architecture is therefore often very hard and costly. Adding a network of microphones is thus generally not an option, and finding a way to record sound with the existing architecture could be beneficial. In the rest of this report, when talking about PA systems, this specific type of system is meant.

2.2. Use cases in public address systems

In a PA system, several use cases exist for recording with loudspeakers.Three key cases are given here:

Figure 2.1: An example of a PA system speaker integrated into the ceiling (indicated in green).

- *Event detection* - In case of emergencies, quick and complete evacuation of an area or building is imperative. Using loudspeakers as microphones, one might be able to detect the presence of people. This way, people, possibly in need of help, can be effectively found. Also, in case of calamities, a recording of the PA system could provide information such as the time of occurrence of a certain event. As a PA system is inherently a security system, this functionality would be an ideal addition to the current systems.

- *Adaptive Announcements* - If during playback of announcements a signal can also be recorded, this recording could be used to give feedback to the system on the announcement. Some examples: if one could hear a train coming in on a platform, the level of the announcement playing over the speakers could be increased to make the announcement more intelligible. An estimation of the room in which the announcement is played could also be made. This can be done by listening to the reflections from a played signal, and making an estimation of the acoustic transfer function (ATF) of a room. This function describes how the signal propagates and reflects in the space, which can then be used for pre-processing of other signals. When evaluating this use case, three possible outputs of the system were defined:

  1. An ambient noise level estimation;
  2. An estimation of the acoustic transfer function of the room;
  3. A full recording (from which the previous two could also be estimated).

- *Espionage* - This thesis will provide certain tools that could be misused for espionage. Due to ethical objections this is not a focus of this thesis. There is, however, no denying that if the audio quality is sufficient, a possible security risk is exposed. Given the long history of research in the audio field, it is also reasonable to assume parties along the way have already tried this, or are actively involved in this use. If from the results from this research espionage seems plausible, a public warning is in place. This is something that needs to be taken into account throughout the project.

  To put the previous statement into perspective, a note needs to be added. Nowadays, people generally have one or more recording devices with them at all time. If someone with malicious intent would try to listen in on a conversation, hacking into someones phone or laptop for a microphone signal would be the logical way to go. This means the possible security risk in the speakers is relatively small compared to other methods of espionage.

The first two use cases were chosen as a focus for this thesis, from which the scoping and research questions will be defined.

2.3. Research objective and research questions

The research objective can be established from the presented use cases.

> *The research objective is to devise a system which can use a loudspeaker to simultaneously play and record audio. This system should consist of components as similar as possible to those in a public address (PA) system.*

The focus of this research will be on combining theoretical concepts with practical insights. The theoretical component exists of building a signal model, a digital signal processing (DSP) element and a conceptual design of an electronics system. The practical component will support the signal model by providing inputs based on experimentation and simulation, which are used for validation and adaptation of the model. To reach the research objective, several research questions are set up and presented below:

RQ 1 - What components does a PA system consist of?

RQ 2 - How can audio be recorded using a loudspeaker?

- **2a -** What is the response of a loudspeaker when used as a recording device?
- **2b -** What is the dynamic range between the playback signal and the recorded signal?
- **2c -** What electronic devices are needed for recording during playback?

RQ 3 - How can signal processing be utilised to increase the quality of the recording?

- **3a -** What is a suitable signal model for this application?
- **3b -** What algorithms can be used to improve the quality?

2.4. Organisation of this thesis

A look at the similarities and differences of the different devices will provide some understanding of the possibilities and pitfalls of using a loudspeaker as a recording device. A primer on loudspeaker and microphone technology is provided in Chapter 3. For those who are experienced in these fields it is recommended to skip this chapter.

Chapter 4 provides an identification of components that make up a PA system to answer 4, and concludes by reducing the system to the most relevant components.

The second research question, which deals with the possibility of recording with a loudspeaker, is discussed in-depth in two parts. **RQ 2a** and **RQ 2b** are discussed in Chapter 5, which details measurements performed in an anechoic room to identify loudspeaker recording characteristics. **RQ 2c** is discussed in Chapter 7. This chapter provides a discussion of an equivalent electronic model of a loudspeaker, which will provide some information for the signal model. After this, a concept-level design for the electronics system is presented to provide simultaneous recording and playback capabilities.

To answer **RQ 3a**, a signal model framework for the system is presented in Chapter 6. This framework covers the system for playback and recording, as well as a digital filter. The filter is described by means of the different desired outputs derived from the use cases described in Section 2.1. Chapter 8 discusses a simulation, performed to answer **RQ 3b**. A simulation is made using the signal model framework, and a simple DSP filter is implemented to prove the feasibility of using DSP to improve the signal quality. This simulation also verifies the feasibility of the entire system, by showing the results that can be retrieved from a system that records audio while playing.

Lastly, Chapter 9 will conclude the thesis with a summary of the presented work, along with a reflection on the work and recommendations for future work to be performed in this area.

3

Loudspeakers and microphones: a technological comparison

As not everyone is familiar with the technologies used in loudspeakers and microphones, a more in-depth evaluation of the technologies in both devices is warranted. This chapter will provide this evaluation, as well as looking at the important similarities and differences between the devices. The focus of this chapter will be solely on moving-coil electrodynamic loudspeakers and dynamic microphones. From this point on, the full nomenclature will be omitted for brevity. Unless noted otherwise, all mentions of loudspeakers will refer to moving-coil electrodynamic loudspeakers, and all mentions of microphones will refer to dynamic microphones.

3.1. Loudspeaker technology

In its simplest form, loudspeakers consist of a cone connected to an electric motor. This motor consists of a coil inside a small air gap of a magnetic pole, called the voice coil. This coil drives the cone to produce the air pressure difference for the sound. The motor and cone are suspended by the spider suspension, which provides most of the stiffness, and the surround, which keeps the cone centred and provides some extra stiffness. Finally, a dust cap shields the motor from dust and debris, and the chassis holds everything together [1]. A schematic cross-section of a loudspeaker is shown in Figure 3.4a. All of these separate components influence the performance of the loudspeaker in some way. These components can be grouped together in some important characteristics.

The **moving mass** is the total mass of the moving parts. Coil wire diameter and length, cone size, and cone material are the most important parameters that influence this mass. The moving mass influences the output sensitivity, as the motor needs to work more to get to the correct excursion (excursion is the movement distance back and forth about the resting position). A way to increase the sensitivity again is by increasing the Bl product, which is the integral of the magnetic field strength over the coil length (usually given in Newton per Ampere as it relates to the force generated by a current). This can be done by for example using a stronger magnet or using a smaller wire diameter, which increases the coil length inside the magnetic field. This in turn, increases the moving mass, so the design of this is always a trade-off. The size and material of the cone depends on the use of the loudspeaker. Generally, a larger cone means a lower frequency response of the loudspeaker. The material is usually chosen to have a certain stiffness. As the vibration in the cone is generated from the voice coil, at higher frequencies the propagation of these waves starts to show in the cone itself, and it will deform. This will result in a poor transfer characteristic and is why the stiffness is important. This deformation is called cone breakup, and is sometimes designed to be a specific way or at a specific point.

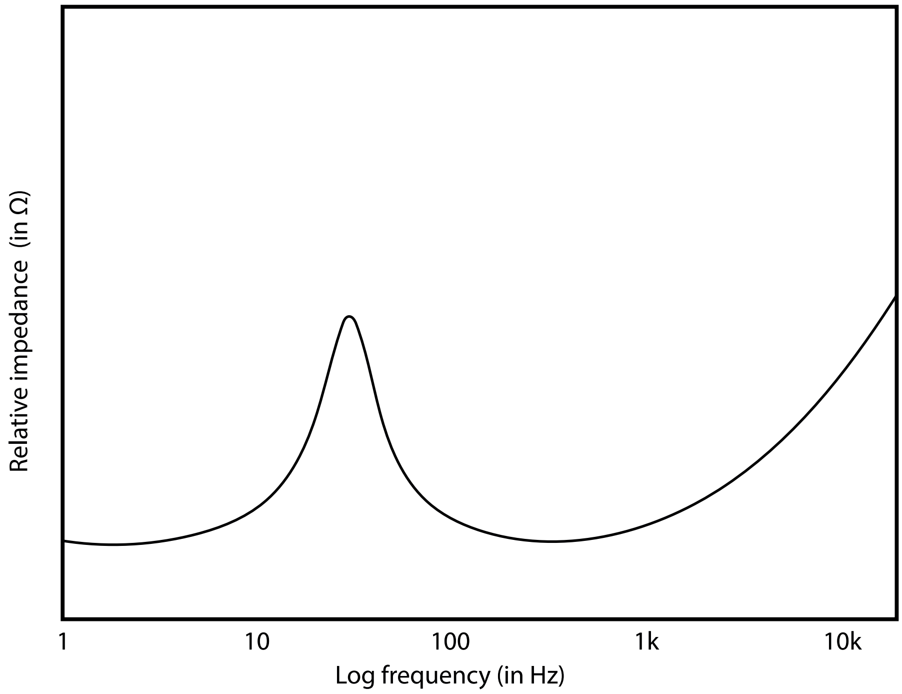


Figure 3.1: A loudspeaker impedance curve. The fundamental resonance of the system can clearly be seen.

Compliance is the reciprocal (inverse) of the stiffness of the suspension. The suspension consists of the spider, close to the voice coil, and the surround, at the edge of the cone. This suspension ensures the voice coil and cone stay centred and return to their neutral position. The spider provides most of the stiffness in this suspension. While the surround provides some additional stiffness, it is mostly responsible for keeping the outside of the cone in the centre. The compliance in the suspension and the moving mass of the system determine the frequency of the 'fundamental resonance' of the loudspeaker. Below this frequency, the system is stiffness controlled, which means most of the energy is needed to deflect the suspension, and relatively little energy is needed to move the moving mass. Above the resonance, the inertia of the mass becomes greater than the stiffness of the suspension and the system becomes mass controlled.

Loudspeakers are usually driven from a voltage amplifier, but the force in the motor is induced by the current (through the Bl product). The **impedance** of the loudspeaker determines the current-to-voltage ratio of the output power of the amplifier. This impedance is related to the mechanism of the speaker as well. Generally the fundamental resonance determines the lower limit of the usable frequency range of the loudspeaker. At the fundamental resonance of the loudspeaker, a small current can produce a relatively large movement. The high impedance at this frequency means the (voltage) amplifier can use its power to produce a high voltage with a relatively low current, and this is very efficient. On top of this, the phase difference between the voltage and current driving the loudspeaker is also 0° at the fundamental resonance. At higher frequencies the inductance of the voice coil starts to affect the impedance and it rises again. An example of an impedance curve for a loudspeaker is shown in Figure 3.1.

3.2. Microphone technology

A dynamic microphone generates its output from the movement of a voice coil inside of a magnetic field, usually produced by a permanent magnet [4]. The movement comes from the diaphragm connected to the coil, which is suspended above the construction and reacts to the incident air pressure. A schematic cross-section of a microphone is shown in Figure 3.4b.

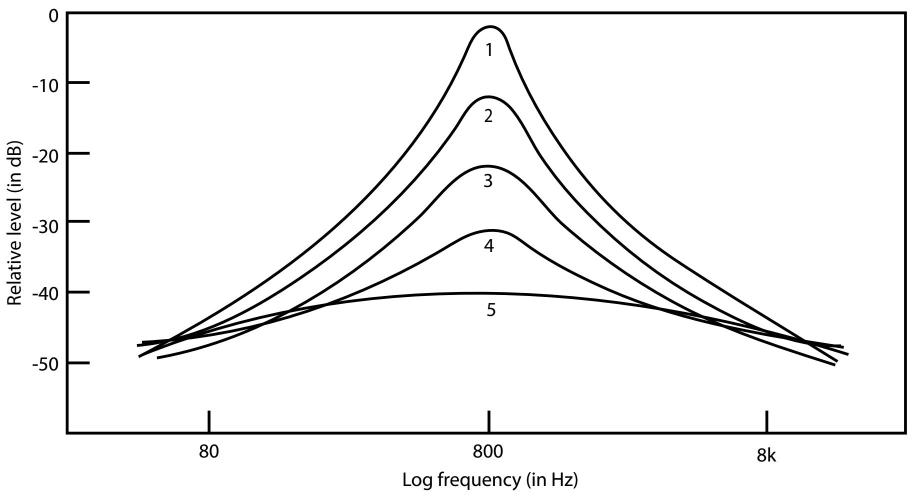


Figure 3.2: A microphone response with an undamped diaphragm (curve 1). Curves 2 to 5 show the curve with an increasing amount of damping.

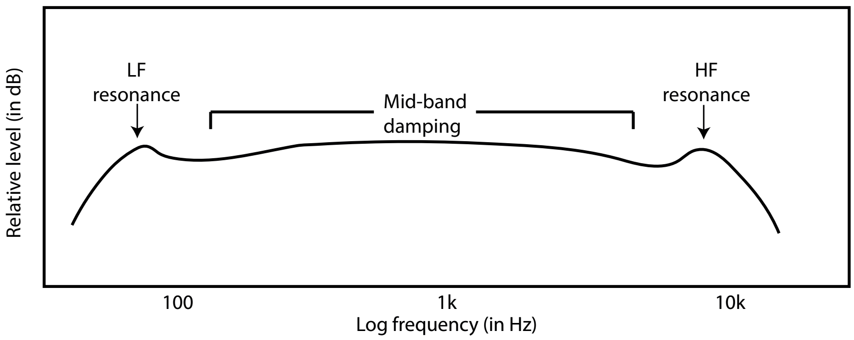


Figure 3.3: An illustration of the net microphone response after compensation.

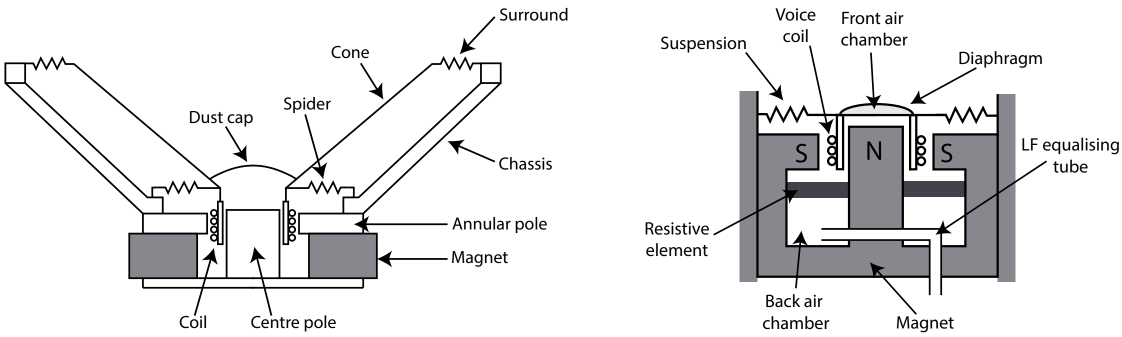
As with loudspeaker, a microphone system has a resonance determined by the mechanical parts of the system. In the case of a microphone, however, this resonance is placed in the geometric mean of the intended frequency range. This resonant peak is then damped to achieve a reasonable flat response over the intended frequency range. This is shown in Figure 3.2. This damping is done by adding a resistive element in the air chamber behind the coil. This means the microphone is resistance controlled over its useful frequency range.

The microphone response drops off quite a lot at the lower and higher frequency range. Compensation is usually employed to boost the response in these frequency ranges. A tube (or hole) in the back air chamber that exits to the outside air allows for low frequency compensation. The dimensions of this tube are chosen such that the air mass inside it resonates with the compliance provided by the back air chamber. This resonance is usually in the range of 40-100 Hz. The high frequency falloff is compensated by creating a small air chamber just inside the diaphragm itself. This chamber resonates as well, usually in the 8-12 kHz range. Figure 3.3 shows how a net microphone response looks after compensation.

To adjust the signal- and device characteristics, dynamic microphones often also contain a transformer inside them. Depending on the sensitivity, the transformer either boosts of lowers the voltage to a suitable range for recording device inputs. Also the transformer isolates the voice coil from DC signals and provides an output impedance in a suitable range for recording.

3.3. Similarities & differences

From Figure 3.4 it is evident that loudspeaker and microphone technology are very similar. Both devices have the same basic configuration, where a coil inside a magnetic field is attached to a diaphragm/cone suspended above it. Also the principles of the loudspeaker motor and the microphone generator are two sides of the same device. In addition to this, both devices utilise a permanent magnet to produce the magnetic field around the coil.



(a) Schematic cross-section of the side view of a loudspeaker (b) Schematic cross-section of the side view of a microphone

Figure 3.4: Schematic cross-sections of a loudspeaker and a microphone. The similarities in the construction can be seen very clearly.

As loudspeakers and microphones have different uses, they are optimised differently for the best intended transfer. This means that while being similar in mechanics, they are not identical. Most notable is probably the size. While microphones have a typical diaphragm radius of roughly 9 mm to 20 mm, moving coil loudspeakers come in sizes varying from roughly 15 mm to 380 mm and even larger in some specific cases.

This size difference leads to a difference in mass. The moving mass in a microphone is usually less than 1 gram. The smallest loudspeakers also occupy this range of moving mass, but the larger loudspeakers exceed this and often range from a few grams to over a hundred grams, depending on size and material. This increase in moving mass is expected to impact the recording transfer of a loudspeaker. With a larger moving mass, the loudspeaker will probably record signals with a lower sensitivity than a microphone. On top of this, the loudspeaker will be more sensitive to lower frequencies compared to higher frequencies, as the fundamental resonance is lower than in a microphone.

Another difference is the nature of the enclosure. Naturally, loudspeaker enclosures are generally much larger than most microphones, but other differences are also present. Loudspeakers usually only enclose the space between the spider and the back, to prevent dust coming in near the coil. Most of the cone moves in free air, and thus the enclosure around the speaker also determines the final transfer. In microphones, the entire system is closed off. Additional elements such as the resistive elements in the back air chamber are implemented to get the desired response before it is enclosed. With the resistive damping from microphones missing in loudspeakers, it is expected that the mechanical resonance of the loudspeaker will produce a large peak in the recording transfer.

Loudspeakers also do not have a resonant air chamber in the diaphragm itself. This means that compared to microphones, they are expected to perform poorly in the high frequency range. Low frequency compensation is most probably irrelevant, as the mechanical resonance of a loudspeaker is at a much lower frequency than the resonance of a microphone.

3.4. Summary

This chapter has provided a technological primer for this thesis. Loudspeaker technology and microphone technology have been introduced to the reader, and from a comparison of these technologies expectations have been set. These expectations are as follows:

- As the governing principle of loudspeakers and microphones are each others reciprocal, loudspeakers are expected to also work as microphones.
- The difference in size and mass is expected to produce a lower sensitivity in recording loudspeakers compared to microphones.
- The lack of resistive damping in loudspeakers is expected to produce a significant peak at the fundamental resonant frequency in the recording response of a loudspeaker.
- The lower resonant frequency of loudspeakers, together with the lack of high frequency recording compensation, is expected to make the loudspeaker more sensitive to lower frequencies.

With the information presented in this chapter, a primer on loudspeaker and microphone technologies is provided to the reader. This primer includes a comparison of the technologies, from which expectations about the research have are set. These will be used for evaluation of the measurements presented in Chapter 5.

4

PA system component analysis

As the research objective explicitly demands similarity to a PA system, identifying the system components in the envisioned environment is essential. To this end, this chapter will provide an answer to the first research question:

RQ 1 *- What components does a PA system consist of?*

The answer to this question is best found by analysing a representative, recent PA system such as the Bosch Security Systems PRAESENSA system [10]. This is the most recent system from Bosch Security Systems at the time of writing. Using this system as a reference, we can identify the current generation of systems in use by the industry, which gives the most accurate view of the components already in place. The key principles are common for the most ordinary PA systems, with the exception of some of its modern features such as DSP capabilities in all call-stations and amplifiers and an added security layer in the communication protocol.

4.1. Full system architecture

The PRAESENSA system is based on digital communication using the Internet Protocol (IP) [11]. Digital communication handles most of the network, up to the amplifiers for the loudspeakers. A schematic representation of a PA system is given in Figure 4.1. The components of the full system are:

- **System Controller**: The system controller manages all system related functions in the system. It routes all the audio from the sources to their destinations on the network. System safety supervision as well as storage and playback of predetermined messages and tones are done by this device. It collects and logs all status information from the connected devices, and reports if there is a fault.
- **Multi-channel amplifiers**: The amplifier receives the digital signal from the system controller, and plays this over lines with multiple speakers connected to them. Multiple output channels allow for driving different zones from a single amplifier, and a spare amplifier channel takes over if one of the other channels fails. Digital signal processing capabilities are integrated per channel, as well as device and line supervision and monitoring.
- **Loudspeakers**: The loudspeakers generate the sound from the signal on the amplifier output. Depending on the system, a range of different loudspeakers can be chosen. Most popular are the ceiling-mounted and cabinet loudspeakers [12, 13].

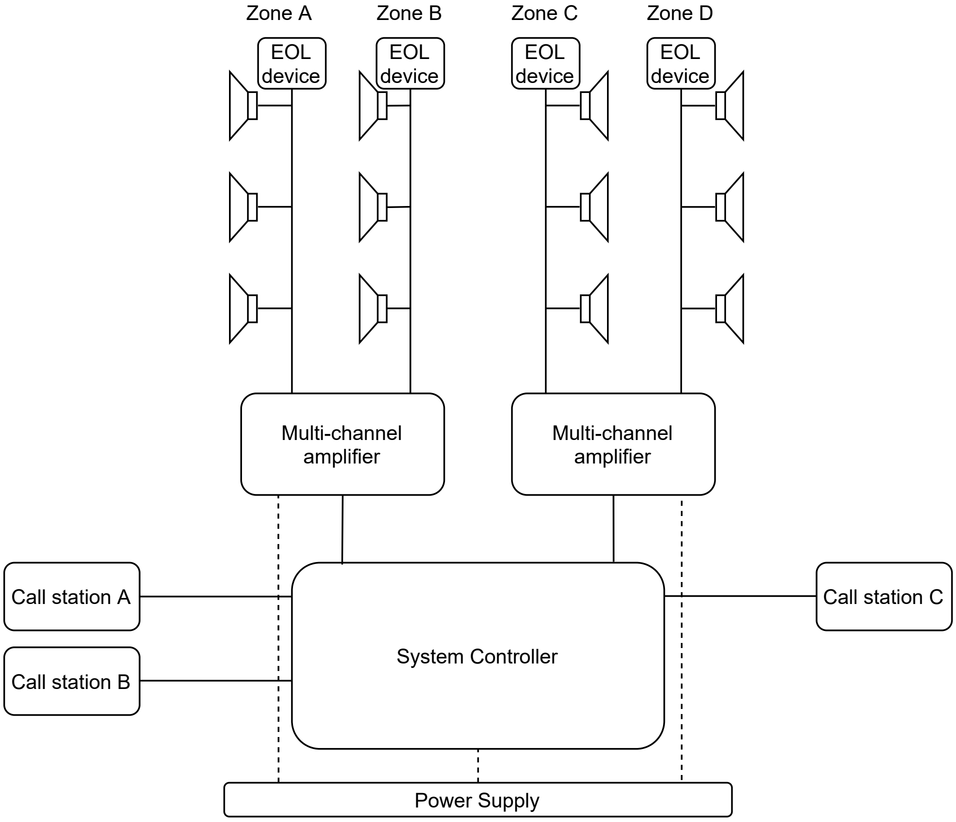


Figure 4.1: A schematic example of a public address system, based on the components of the Bosch Security Systems PRAESENSA system. The system controller handles the communication between the separate components. The amplifiers have multiple output channels to drive different zones in a building, and multiple loudspeakers are usually connected to each of those channels. The end-of-line (EOL) devices reflect a safety signal back to the amplifier to signal the line is still intact.

- **Power Supply**: The power supply is an external unit that powers multiple devices in the PA system. It takes the AC supply voltage from the outlet and converts it into the DC voltage required for each of the PA system components. Depending on the system configuration the unit can power multiple components from a single unit (centralised), or be mounted with the separate components (decentralised) to reduce cabling cost. A backup battery is integrated, to ensure the system stays operational during power outages.

- **Call Stations**: These are the stations used for live communication over the system. They allow the user to select the destination, and either talk into a microphone or send a signal from an external audio input. These devices serve as audio inputs for the system controller, while also providing system feedback to the user. A basic call station consists of at least a microphone, a monitoring speaker and an audio input with device monitoring, but the PRAESENSA call stations also include a touch screen for visual feedback to the user.

- **End-of-line (EOL) devices**: These are small devices connected to the end of the loudspeaker line, to increase the reliability of the loudspeaker line integrity supervision. This device communicates with the amplifier using a high frequency pilot tone. If the amplifier channel does not receive the signal from the EOL device, this means the line is broken or a loudspeaker is disconnected along the line.

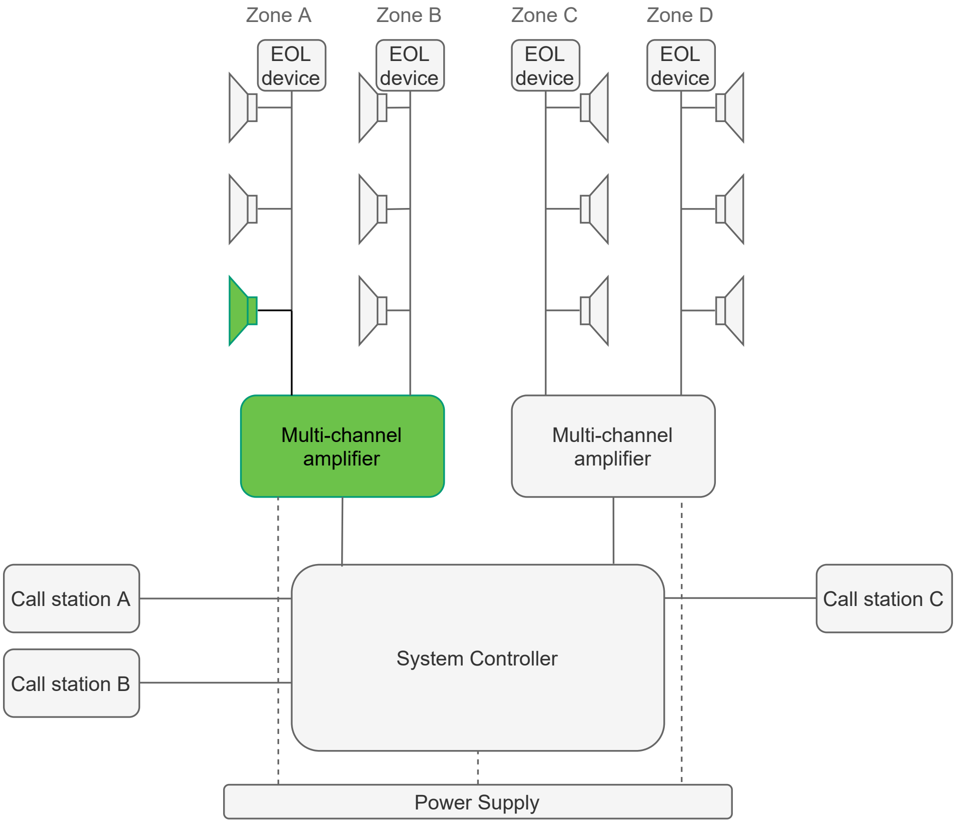


Figure 4.2: The PA system diagram, with the key relevant components for this thesis highlighted in green.

4.2. Simplifying the system

For this research the key components to consider are the amplifiers and the loudspeakers. With the DSP capabilities present in the amplifier, integration of signal processing algorithms is possible without the need of the system controller. The monitoring capabilities together with the EOL devices offer some insight as well, as they show us the amplifier has two-way communication over the loudspeaker line. The other elements such as the call stations, system controller and the power supply do not have an impact on this research, as there are no variables that can typically be controlled that will impact the results of this study. Therefore these components are left out.

PA systems typically have a large number of loudspeakers connected to a single output channel, which complicates the task of measuring the signal from a single loudspeaker independently. Therefore to isolate this problem, a single loudspeaker system is chosen for analysis of the research questions.

With the system analysis presented in this chapter, research question **RQ 1** is answered. The Bosch Security Systems PRAESENSA system is chosen for evaluation, and the system architecture is described. As most of this system uses IP communication, the system can be simplified to the key components. A system of one amplifier and one loudspeaker will provide the analogue signal path, as well as DSP capabilities. The key system components that need to be evaluated are shown in Figure 4.2.

5

Loudspeaker measurements and parameter identification

This chapter describes the characterisation off several loudspeakers. The goal of this characterisation is to answer two parts of research question 2:

RQ 2a *- What is the response of a loudspeaker when used as a recording device?*
RQ 2b *- What is the dynamic range between the playback signal and the recorded signal?*

The answers to these research questions will provide essential information on the recording process. The dynamic range is pivotal in achieving a successful recording, as this will determine if the recorded signal is picked-up by an analogue-to-digital converter or if it is completely masked by the playback signal. The answer to **RQ 2b** is used in Chapter 7 to evaluate the necessary electronic components in the system. The answer to **RQ 2a** is combined with the signal model from Chapter 6 in the simulation in Chapter 8, and it can be used to equalise the recorded signal in the DSP.

5.1. Loudspeaker measurements

To assess the performance of a loudspeaker as a recording device, it is essential to describe the measured signal in a consistent framework such that these can be compared. This section will discuss a number of quantitative characteristics that are used and their respective methodologies for acquiring the data.

5.1.1. Required characteristics

First the required characterisations will need to be determined. Three of the most relevant characterisations are [4]:

1. **Frequency response along the principal axis** Measuring the frequency response allows one to compensate for this frequency response to obtain a recording with a relatively flat response.

2. **Directional response** As not all signals are on axis with the device, it is useful to know how the recording response changes for different incident angles.

3. **Output sensitivity** The output sensitivity is a unit of the electrical signal amplitude of a microphone for a certain sound pressure level (SPL), given in mV/Pa. This will allow for a quantitative analysis of the system as this can be used to calculate the signal level of the recording.

5.1.2. Measurement methodology

Environment

The measurements described in this section were done in a full anechoic room at the acoustic lab of Bosch Security Systems B.V. located in Eindhoven. Removing the room reflections allows for a more precise measurement of the device under test (DUT). Temperature and air pressure also affect the measurements [1], so there needs to be a compensation for these effects. This can be done with a reference microphone and will be discussed further on in this section.

Device under test (DUT)

In all cases discussed here the DUT is a loudspeaker from the Bosch LC1 ceiling loudspeaker range [12]. This range of loudspeakers is representative of the average ceiling loudspeaker in a lot of PA applications, and the 6W models from this range are among the most popular of the whole loudspeaker portfolio. The range includes five different types of speakers: WM06E8, WC06E8, UM06E8, UM12E8

Figure 5.1: Image of an LC1-WM06E8 loudspeaker range [12]

and UM24E8. The numbers in the names indicate the rated power of the loudspeakers (6W/12W/24W). Of the WM06E8 and UM06E8 two samples were available, and one sample each was available of the other types. As one can connect the speakers to different taps in the transformer for different output powers, it was decided to perform two more measurements on the first WM06E8 sample on the taps for 3W and 1.5W. All the other measurements were performed using the 6W taps on the transformer. In total nine measurements were made:

1. WM06E8 #1 at 6W
2. WM06E8 #1 at 3W
3. WM06E8 #1 at 1.5W
4. WM06E8 #2 at 6W
5. WC06E8 at 6W
6. UM06E8 #1 at 6W
7. UM06E8 #2 at 6W
8. UM12E8 at 6W
9. UM24E8 at 6W

Figure 5.2: A close-up of the setup used for measuring the recording response of loudspeakers. The height of the loudspeaker did not match that of the microphone and the device under test (DUT). Also the microphone position is not ideal, as it should be right next to the DUT. These non-idealities should be compensated for in the measurements.

Tools and setup

Suspended in the anechoic chamber are a loudspeaker and a reference microphone. The loudspeaker is placed in a spherical enclosure, as to reduce the diffractions that can happen at the edges of the baffle [14]. The reference microphone is suspended in front of the loudspeaker. Behind the microphone, the DUT is held up by a stand. The DUT is placed inside a wooden plate of 60 cm x 60 cm, the size of a standard ceiling tile. The DUT is placed in a LC1-MFD fire dome. This is how the DUT would normally be placed inside a building, with a fire dome or acoustic cap closing off the back and inside of a ceiling tile. A picture of the setup is shown in Figure 5.2.

It is beneficial to have the reference microphone as close as possible to the DUT to ensure proper calibration. Because of the size of the DUT and the mounting panel, the reference microphone could not be placed near the membrane, but instead had to be placed in front (15 cm). This will mean the level calibration will be off. This needs to be compensated for in the results.

Due to the construction in the room, the DUT and the reference microphone could not be placed perfectly horizontal with the loudspeaker. This means the signal received at the microphone and loudspeaker is not the perfect on-axis signal of the loudspeaker, and this could have some impact on the calibration and measurements. The stand on which the DUT is mounted can be turned to specific angles via a motor controller outside the chamber, which allows for measurements in quick succession. As the turning angle was horizontal, the height difference is not altered by turning the DUT.

Calibration & compensation

To compensate for the temperature and air pressure in the room, the measurement needs to be calibrated. The reference microphone and loudspeaker have a known transfer, which is already compensated for by the software. Using a software tool, repeated measurements were taken with the reference microphone while playing white noise over the loudspeaker. Between these measurements, the equaliser on the output channel was adjusted repeatedly to get the frequency response of the measurement as flat as possible. In the range of 100 Hz to 20 kHz the frequency spectrum after calibration was flat to within 0.01 dB. With this, the inaccuracy due to the room conditions has been reduced to a negligible amount. Ideally this calibration would be done for each new measurement, or at least every time the DUT is changed, but because of time constraints it was only done once at the beginning of the session.

As described in the 'Tools and setup' section, some height and distance effects need to be compensated for in the end results. The sensitivity measurements are performed at 94 db SPL, measured at the calibration microphone (SPL_m). To determine the SPL at the DUT (SPL_l), the distance between the DUT and the microphone needs to be known. The horizontal distance between these devices is 15 cm. Also, the centre of the DUT has a vertical distance of 6cm to the centre of the loudspeaker, and a horizontal distance of 50 cm. Using simple geometry, the total distances are calculated to be 50.36 cm from loudspeaker to DUT (d_l), and 35.11 cm from loudspeaker to reference microphone (d_m). The resulting SPL at the loudspeaker can then be calculated as:

$$\mathrm{SPL}_l = \mathrm{SPL}_m - 20log_{10}(\frac{d_l}{d_m}) = 94 - 20log_{10}(\frac{50.36}{35.11}) = 90.87dB \tag{5.1}$$

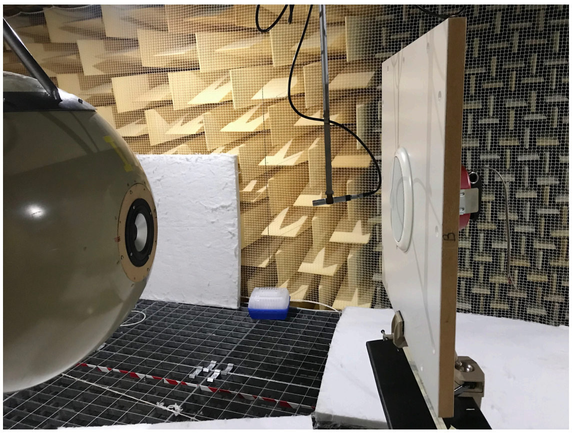
So the sensitivity measurement is done at 3.13 dB lower then usual. For the frequency responses, which are normalised using the sensitivity, this does not influence the shape of the response because it is simply a bias over the entire range. Only when relating the output of the DUT to the incoming SPL this bias has to be accounted for. As the sensitivity is given in mV/Pa (94 dB SPL $= 1$ Pa), the sensitivity needs to be scaled by a factor q:

$$q = \frac{1}{(P_0)10^{\frac{\mathrm{SPL}_l}{20}}} = \frac{1}{(20 * 10^{-6})10^{\frac{90.87}{20}}} = \frac{1}{0.70} = 1.43 \tag{5.2}$$

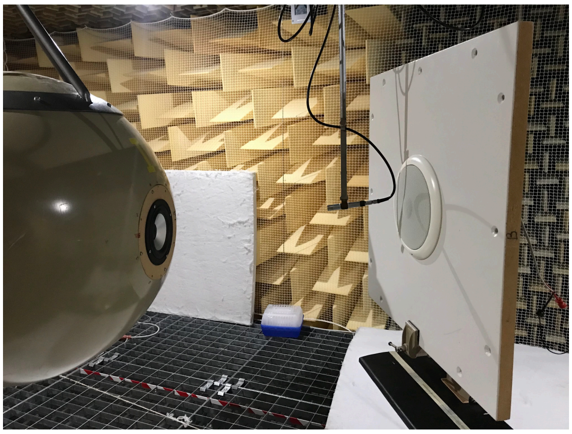
Where the reference pressure $P_0 = 20\mu Pa$ for air is used [1].

Measurements

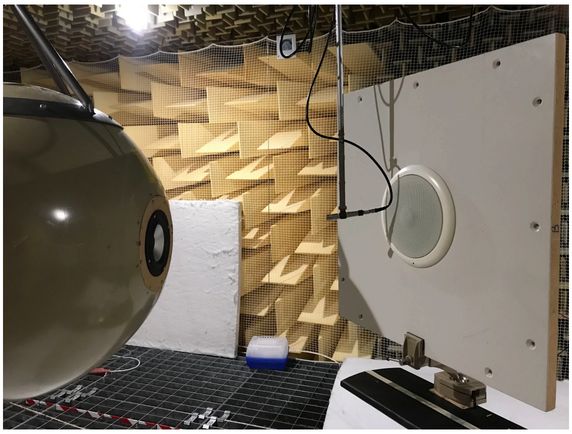
Each measurement consists of two parts: a frequency sweep measurement and a 1kHz sine measurement. The measurement at 1 kHz is used for determining the output sensitivity of the DUT, as well as normalising the frequency measurement. After this, the stand holding the DUT is turned 15°. The measurement is repeated in these increments until the DUT is at a 75° angle. Ideally the measurement would go up to 90°, but due to the construction this is not possible. As the DUT is assumed to be radially symmetric along the primary axis, only one quadrant was measured. Figure 5.3 shows the setup in the different measurement angles.



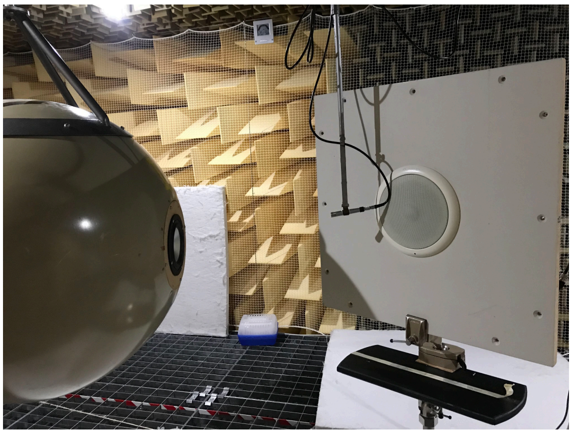
(a) The DUT at 0°



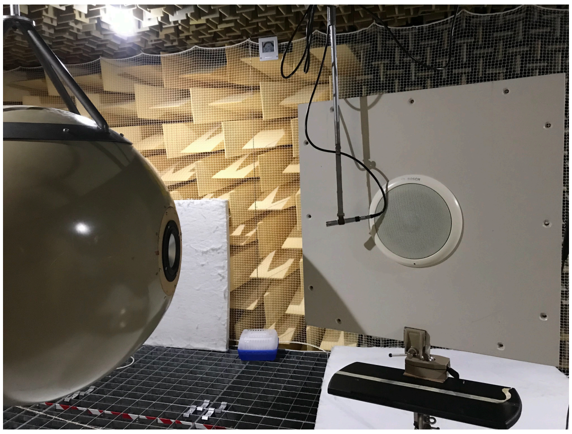
(b) The DUT at 15°



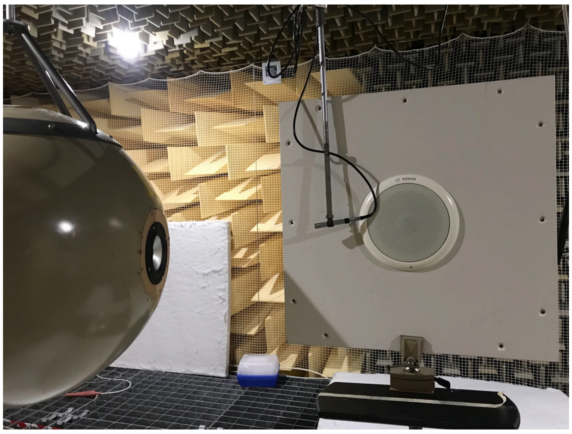
(c) The DUT at 30°



(d) The DUT at 45°



(e) The DUT at 60°



(f) The DUT at 75°

Figure 5.3: The angles at which the DUT was measured. As the device is assumed to be radially symmetric along the primary axis, only one quadrant was measured. Because of the mounting plate and the microphone suspension, a 90°angle could not be measured.

5.2. Results & analysis

The resulting frequency and sensitivity plots are shown in Figure 5.4, Figure 5.6 and Figure 5.7. For brevity the results are only shown for a single device type. The remaining measurement results are shown in Appendix B.

5.2.1. Frequency response

In Figure 5.4, as well as in Appendix B.1, the measured recording responses of the DUTs are shown, together with their playback response as specified by the manufacturer. Comparing these responses, two main trends can be seen across the whole speaker range:

- Between roughly 150Hz and 750Hz a significant increase in the response is seen. This is most probably the result of the DUTs resonant frequency. As shown in Chapter 3, the resonance of microphones is usually significantly damped. The large peak in the response is expected as described in Section 3.4. Using the fire dome to mount the DUT means it has a small enclosure, which causes the resonant frequency of the system to be higher than the free air resonant frequency of the driver [1]. The peak value of $20dB$ is significant, and compensation can be useful to ensure signals in this frequency range do not mask other signals in the higher frequency range.

- From 5kHz upwards, the recording response starts to drop off significantly. This is most likely due to the mass of the diaphragm being larger than it is in a microphone. In the high frequency range the recording response is determined mostly by the mass of the diaphragm [15], and no high-frequency compensation is present. In all recording responses a slight bump of a few dB can be seen around 5kHz before the response drops off. This could be due to a resonances in the cone itself, similar to the air-chamber in microphone diaphragms. As discussed in Section 3.4 the lack of high frequency response was expected.

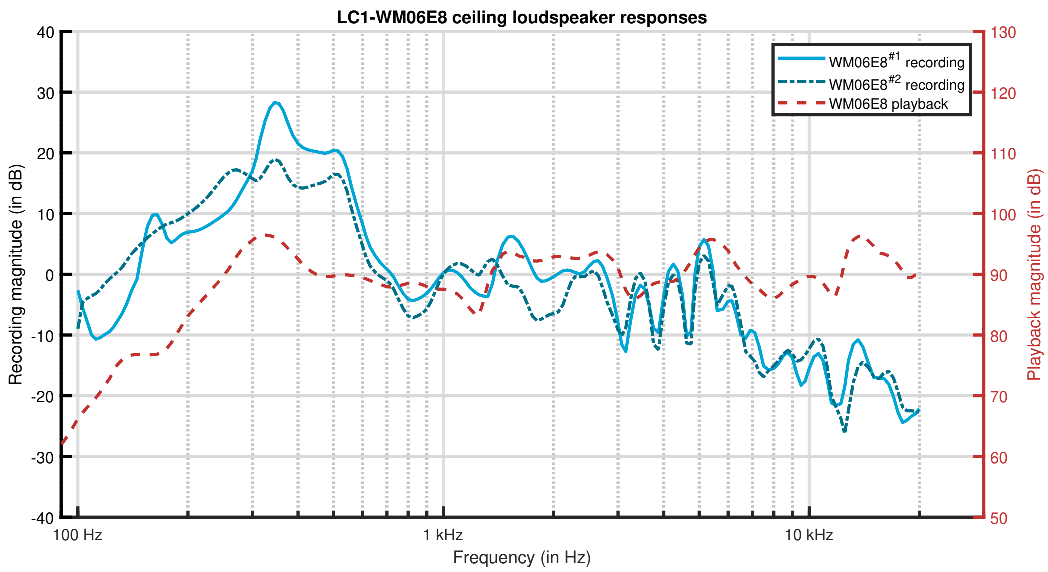
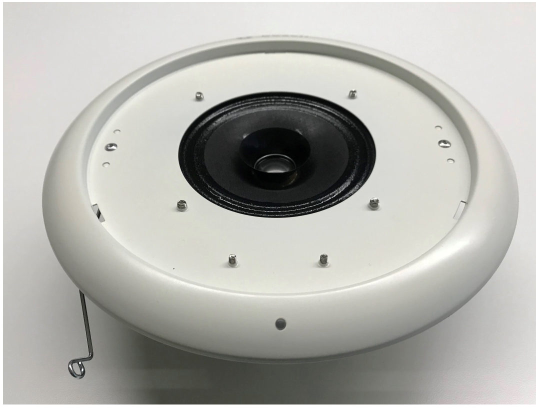


Figure 5.4: The on-axis measured recording response of the WM06E8 loudspeaker, compared with the playback response. The blue solid and dash-dotted curves represent the recording response of two different loudspeakers. The red dashed curve is the playback response, as given by the datasheet.

5.2.2. Directional response

The directional response is shown in Figure 5.6 and Figure 5.7. Figure 5.6 shows the full frequency responses for 4 different angles, while Figure 5.7 shows only the sensitivity at 1kHz in a polar plot. From these figures, two significant observations can be made. The responses are similar to a large extent, apart from a shift in level. Potential compensation for the loudspeaker recording response will then perform roughly the same for all incident angles. Another interesting observation is the sensitivity at a $45°$ angle roughly doubles compared to the $0°$ degree sensitivity. As this is not on axis with the coil movement, an increase of the sensitivity compared to a $0°$ degree angle seems unlikely. In an attempt to identify where this increase comes from, the grills from two loudspeakers (WM06E8 and UM06E8) was removed to look at the drivers. These are shown in Figure 5.5a and Figure 5.5b. On both of these drivers a whizzer cone is attached [16]. These are small, stiff attachments to a cone to improve high frequency performance of loudspeakers. These attachments are mounted directly to the coil-cone interface. As the angle on these whizzer cones is 45 degrees, these could be a possible cause for the increase in sensitivity, although this was not tested.

(a) An LC1-UM06E8 loudspeaker with the grill removed.



(b) An LC1-WM06E8 loudspeaker with the grill removed.

Figure 5.5: Two loudspeakers from the LC1 range with their grill removed. In the centre of the cone a small whizzer cone is visible.

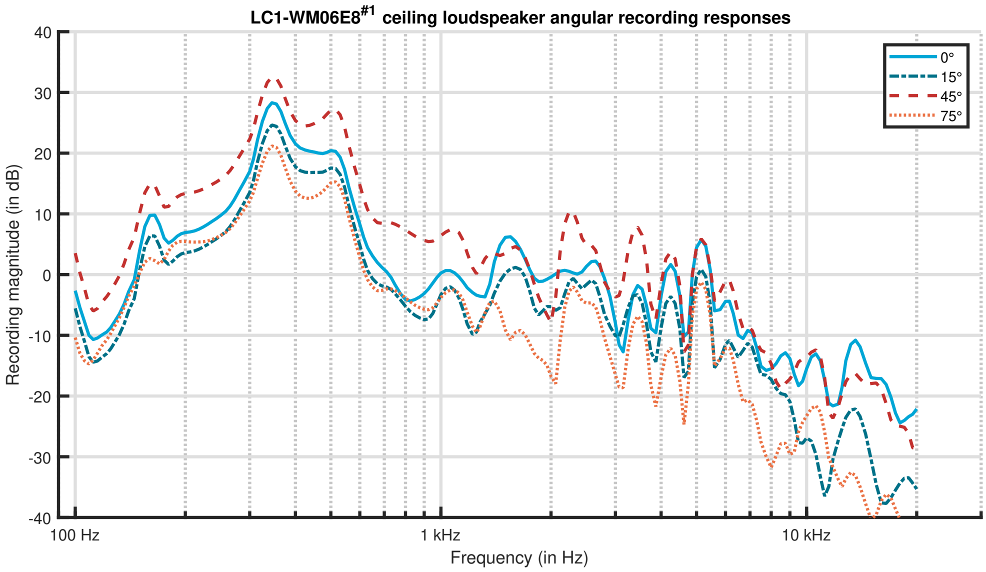


Figure 5.6: The measured frequency response of the first WM06E8 loudspeaker as a microphone over 4 different angles ($0°$,$15°$,$45°$,$75°$).

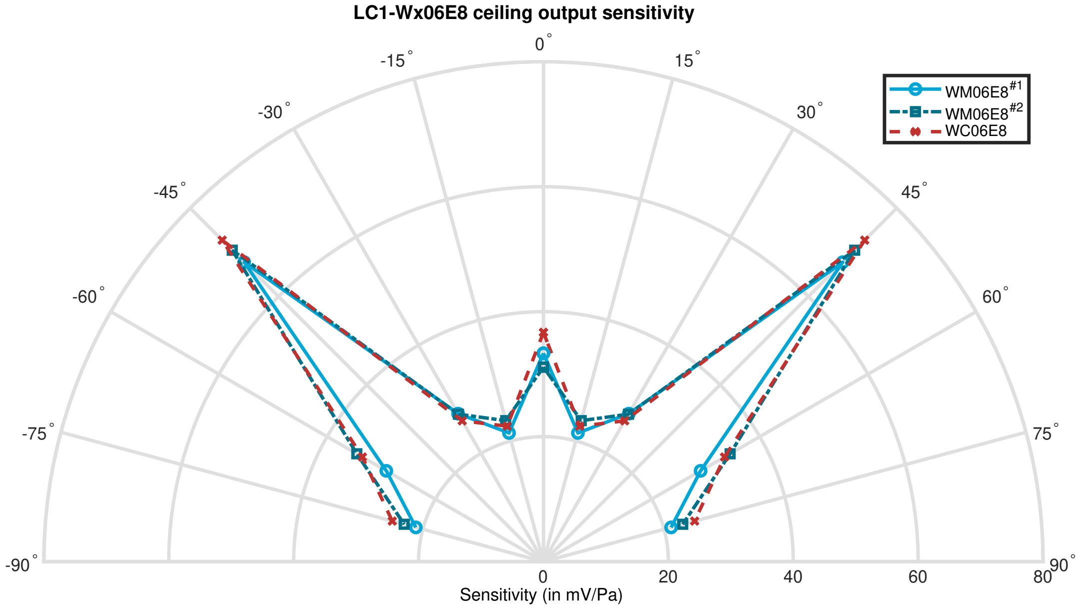


Figure 5.7: Polar sensitivity plot of 3 loudspeakers as microphones (2x WM06E8 and 1x WC06E8) as seen at the transformer connection. Interesting to note here is the increased sensitivity at 45°. The plot was generated using data from one quadrant, mirrored on the 0° axis. The values have been compensated with the factor q calculated in Equation (5.2)

5.2.3. Output sensitivity and dynamic range

The on-axis recording output sensitivity for all the measured loudspeakers is given in Table 5.1. This output sensitivity was measured at the transformer terminals. To compare the recording signal with the playback signal, both signals are referenced to the speaker terminals after the transformer. The referenced output sensitivity is also given in Table 5.1. The last column in Table 5.1 is the calculated dynamic range. As the playback signal also occupies the line, the recorded signal could be masked by the playback signal, as it is much louder. The playback signal on the line is at $100V_{\mathrm{RMS}}$. This is referenced by the same factor of the transformer, based on the selected power tap on the transformer [17]. To determine the dynamic range between these signals, a noise floor needs to be determined. This noise floor is the level at which the intrinsic system noise is masking the actual signal. Anything signal below this level is considered lost. For this application, the noise floor was chosen at $50dB$, the level of a conversation at two to three meters. This level was chosen as it is sufficiently loud above room noise, such as air conditioners. The recorded signal strength was calculated using the difference in level between the noise floor and the sensitivity measurement. This is then compared to the signal strength of the playback signal, and the dynamic range is calculated from this ratio.

DUT (transformer tap)	Sensitivity (in mV/Pa)	Referenced sensitivity (in mV/Pa)	Dynamic range (in dB)
WM06E8 #1 (6W)	23.3	1.24	116.7
WM06E8 #1 (3W)	33.3	1.24	113.5
WM06E8 #1 (1.5W)	47.5	1.24	110.5
WM06E8 #2 (6W)	21.8	1.15	117.2
WC06E8 (6W)	25.6	1.36	115.8
UM06E8 #1 (6W)	17.5	0.93	119.1
UM06E8 #2 (6W)	17.6	0.93	119.1
UM12E8 (6W)	21.9	1.16	117.2
UM24E8 (6W)	21.5	1.14	117.4

Table 5.1: The measurement results for the on-axis sensitivity of the different loudspeakers. The "Referenced sensitivity" column is the sensitivity referenced to the loudspeaker terminals instead of the transformer taps. The "Dynamic range" column is the calculated dynamic range between the recording noise floor and the playback signal. The worst case is using the UM06E8 loudspeakers, where the recording noise floor is 119.1 dB lower than the playback signal on the terminals.

5.3. Summary

This chapter has described the conducted measurements for the identification of the recording loudspeaker characteristics. The LC1 ceiling loudspeaker product range from Bosch Security Systems B.V. was measured in an anechoic room. With the presented results, the research questions **RQ 2a** and **RQ 2b** have been answered:

- The recording frequency response was shown to have a similar shape across the entire product line. Between 150Hz and 750Hz a large increase in sensitivity is present, caused by the resonant frequency of the loudspeaker. A high frequency roll-off is present from 5kHz and up, due to a lack of high frequency compensation.
- The on-axis output sensitivity of the speakers ranges from 0.93-1.36 mV/Pa. At a noise floor of 50dB for recording, this results in a dynamic range of 110.5-119.1 dB for the recorded signal. This means the softest signal recorded from the room is at most 119.1dB softer than the playback signal on the line. Compared to the on-axis recording response of the loudspeaker, a signal with an incident angle of 45 degrees has shown to provide a significant increase in sensitivity (roughly double the sensitivity).

This thesis continues with the signal model of the recording system, in which the characterisations presented here are used.

6

Signal model of the recording system

This chapter will provide an answer to research question 3a:

RQ 3a *- What is a suitable signal model for this application?*

To this end, a more in-depth analysis of the key system components identified in Chapter 4 is provided. The answer to **RQ 3a** can be used as a reference for the implementation of signal processing algorithms, and is used as such in the simulation described in Chapter 8.

6.1. Analysing the system components

In this section, the signal model will be derived with help of a block diagram. To this end, all the individual components of the amplifier-loudspeaker subsystem were identified. The block diagram derived from this is given in Figure 6.1. The blocks inside the dashed rectangle represent everything that are not part of the digital signal processing (DSP). Outside of the dashed rectangle, the blocks represent data or processing in the digital domain. Note that the noise introduced in the different blocks is not shown in this block diagram for readability. Four other simplifications can be seen in the block diagram: the digital-to-analogue converter (DAC) and amplifier are simplified to one block. This simplification is reasonable, as digital-input class D amplifiers are used in the PRAESENSA system. These amplifiers generate a fixed-level pulse train where the width of these pulses is modulated. After filtering out the high-frequency content introduced by the switching, the audio signal information contained in the pulses is left. The switching nature of these amplifiers allows for a digital input, which essentially means the DAC is integrated in the device. The line is also assumed ideal, so it is combined with the amplifier transfer function. As stated in Section 4.2 the system considered consists of one amplifier channel and one loudspeaker. In these cases, the effect of the line is significantly smaller than in the multiple speaker case. This is why the line is is considered to be ideal here. The third simplification is the combination of the line and the analogue-to-digital converter (ADC) at the recording interface. Again, because the line is assumed to be ideal, it can be easily combined with the transfer of the ADC. In the ADC transfer, any analogue circuitry used to condition the signal for the ADC input can be included as well. Lastly, an assumption is made that the movement of the loudspeaker from signal playback will not interfere with the recording process. This assumption is discussed in Chapter 7, where the electronic model of a loudspeaker is presented.

As the signals described here occupy both the digital and analogue domain, the indexing of signals and transfers switches from frequency ω to a frequency bin index k. Taking N equally spaced samples on the frequency range $[0, F_s)$, with F_s the sampling frequency, we can describe the frequency bin ω_k

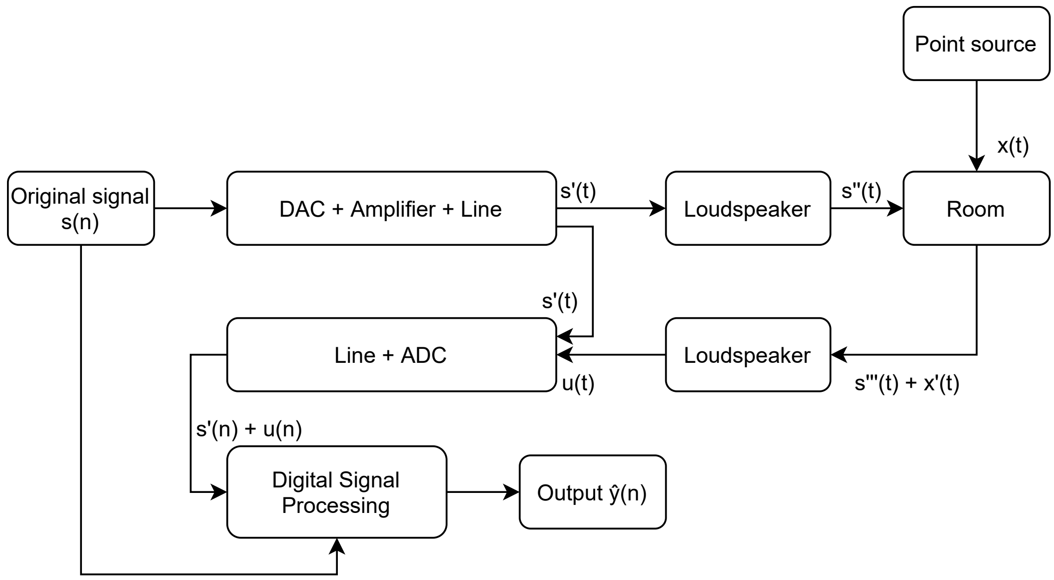


Figure 6.1: Block diagram of the system in which a loudspeaker plays and records sound simultaneously. The different versions of the signals $s(n)$ and $x(t)$ are indicated with accents, denoting alterations to the signal from the different system components. The output $\hat{y}(n)$ is an estimate of the desired output $y(n)$, described in Section 6.2. The noise introduced by the different system components is not included in this diagram.

as

$$\omega_k = \frac{F_s}{N}k, \qquad k = 0, ..., N-1. \tag{6.1}$$

To distinguish the different domains, the notation is slightly abused here and the bin index k is used for digital signals instead of ω_k.

6.1.1. DAC and amplifier

In current audio systems, class D audio amplifiers are often used. This is also the case for the PRAESENSA system described in Chapter 4. These amplifiers are based on generating a pulse-width modulated signal, which is then filtered to remove the switching artefacts from the signal. This leaves a continuous time signal that can be sent to the loudspeakers. The digital nature of these amplifiers means they can be driven from a digital signal, so they act as both an amplifier and a DAC. The amplifier can be modelled as a transfer function which has a digital input and an analogue output

$$S'(\omega) = H_a(\omega)S[k], \tag{6.2}$$

As the amplifier functions as a DAC as well, the transfer function will contain some form of DAC transfer as well as an amplifier transfer. Taking the ideal DAC transfer $D(\omega)$ [18], and the amplifier with the asymptotic gain model described in [19], the transfer will be

$$H_a(\omega) = D(\omega)A_f(\omega), \tag{6.3}$$

where $A_f(\omega)$ is the asymptotic gain model

$$A_f(\omega) = A_i(\omega)\frac{-L(\omega)}{1-L(\omega)}, \tag{6.4}$$

consisting of the asymptotic gain $A_i(\omega)$ and the loopgain $L(\omega)$ [19]. The DAC transfer $D(\omega)$ is

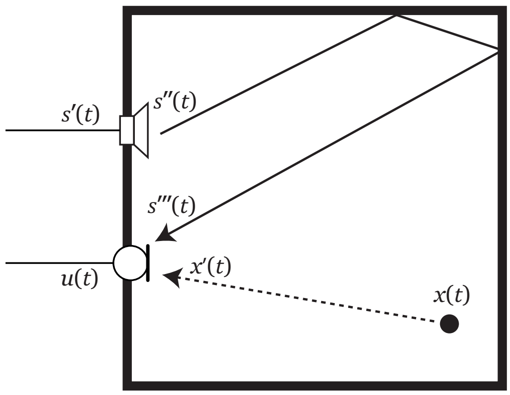
$$D(\omega) = \begin{cases} T, & \text{if } |\omega| \leq \frac{\pi}{T} = \pi F_s \\ 0, & \text{if } |\omega| \geq \frac{\pi}{T}, \end{cases} \tag{6.5}$$

With $T = 1/F_s$ the sampling interval and F_s the sampling frequency. Equation (6.4) and (6.5) serve as an indication of the system. If one were to further specify this component transfer, one could take a different DAC model that includes non-ideal components such as clock jitter and skew. Equation (6.4) is a model used in structured electronic design of amplifiers, where it is used to design the (frequency dependent) gain of an amplifier. Further non-linear characteristics can also be included for a more accurate model. As an addition to Equation (6.2), uncorrelated additive noise is assumed:

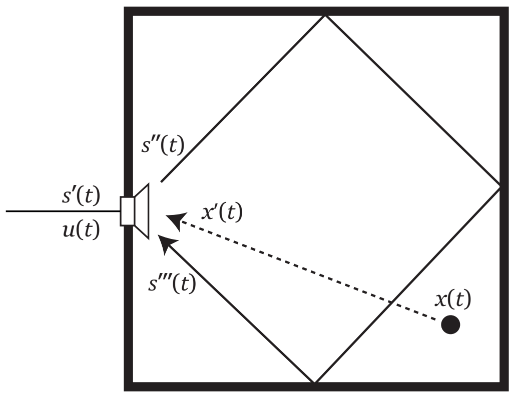
$$S'(\omega) = H_a(\omega)S[k] + N_a(\omega). \tag{6.6}$$

The amplifier model described serves an illustrative purpose, and is not elaborated further in this report. If one were to choose a different or more specific model, substitute the chosen transfer function and noise characteristics in Equation (6.6).

6.1.2. Loudspeaker, room transfer and other present signals



(a) Loudspeaker and microphone are separate devices



(b) Case from this thesis, where the loudspeaker and the microphone are one device

Figure 6.2: Schematic drawings of the use case: a loudspeaker plays a signal and a microphone records what happens in the room. In these figures, only one path per source is drawn for clarity, but the ATF in the models is assumed to contain all paths. The dashed arrow indicates the signal from the point source is uncorrelated to the signal from the loudspeaker. The playback signal is represented by different alterations of $s(t)$. $s'(t)$ represents the amplifier output, $s''(t)$ the loudspeaker output, and $s'''(t)$ represents the signal as it arrives at the recording device. $x(t)$ and $x'(t)$ represent the signal from the point source and how it arrives at the recording device. Finally, $u(t)$ represents the signal output from the recording device.

For the loudspeaker, the room, and the other present signals, a joint analysis before going into detail about the individual components is beneficial. As the goal is to play a signal and record a signal at the same time, we can start with the case where there is a loudspeaker and a microphone together in a room, and then adapt this to the case from this thesis, where the loudspeaker also functions as the microphone.

General case

For the loudspeaker, the room, and the other present signals, it would make sense to look at them together before going into detail about the individual blocks. As the goal is to play a signal and record a signal at the same time, we can start with the case where there is a loudspeaker and a microphone together in a room. This is shown in Figure 6.2a. As the loudspeaker has a specific transfer, the signal $S''(\omega)$ it outputs can be modelled as a filtered form of the amplifier out $S'(\omega)$, with an additive noise source $N_l(\omega)$:

$$S''(\omega) = H_l(\omega)S'(\omega) + N_l(\omega). \tag{6.7}$$

As the microphone picks up the signal in the room, we model the signal in the room with an acoustic transfer function (ATF) [20]. This ATF describes all paths from the loudspeaker to the microphone in the room, including reflections. The signal $S'''(\omega)$ as it arrives at the microphone is

$$S'''(\omega) = H_r(\mathbf{X_l}, \mathbf{X_m}, \omega)S''(\omega), \tag{6.8}$$

where $H_r(\mathbf{X_l}, \mathbf{X_m}, \omega)$ is the acoustic transfer function, and $\mathbf{X_l} = [x_l, y_l, z_l]$ and $\mathbf{X_m} = [x_m, y_m, z_m]$ are the position vectors of the loudspeaker and microphone, respectively. Another signal in the room is modelled as a point source generating a signal $x(t)$. The path from this point source to the microphone, including reflections, is modelled with a different ATF. The point source signal as it arrives at the microphone is

$$X'(\omega) = H_r(\mathbf{X_x}, \mathbf{X_m}, \omega)X(\omega), \tag{6.9}$$

Where $\mathbf{X_x} = [x_x, y_x, z_x]$ is the position vector of the point source. The angles at which $s'''(t)$ and $x(t)$ hit the microphone are dependent on the paths in the room. Assuming the direct path is dominant, the directional microphone response needs to be taken into account. To this end, the microphone response is modelled as dependent on the incident angle θ of the respective signals, and is given as $H_m(\omega, \theta)$. Potential noise in the room is modelled as an additive noise source $N_r(\omega)$. The microphone will also add noise $N_m(\omega)$, and this too is assumed additive. The microphone output $U(\omega)$ can then be written as

$$U(\omega) = H_m(\omega, \theta_s)S'''(\omega) + H_m(\omega, \theta_x)X'(\omega) + H_m(\omega, \theta_n)N_r(\omega) + N_m(\omega). \tag{6.10}$$

Combining equations (6.7), (6.8), (6.9) and (6.10) and omitting the frequency index for brevity, the microphone output is

$$U = H_m(\theta_s)\big(H_r(\mathbf{X_l}, \mathbf{X_m})(H_l S' + N_l)\big) + H_m(\theta_x)(H_r(\mathbf{X_x}, \mathbf{X_m})X) + H_m(\theta_n)N_r + N_m. \tag{6.11}$$

Special case

Using only a loudspeaker for both playing and recording results in a special case of the signal model described above. This case is shown in Figure 6.2b. The majority of the signal model is unchanged, but there are two exceptions. As the line carrying the recorded signal is the same line that carries the signal for playback, this playback signal is also present in the recorded signal. So we extend Equation (6.11) to include this:

$$U = H_0 S' + H_m(\theta_s)\big(H'_r(\mathbf{X_l}, \mathbf{X_l})(H_l S' + N_l)\big) + H_m(\theta_x)(H_r(\mathbf{X_x}, \mathbf{X_l})X) + H_m(\theta_n)N_r + N_m, \tag{6.12}$$

where H_0 is the transfer from the line to the line. In this report, this is assumed to be the trivial transfer $H_0(\omega) = 1$. The rest of the signal model remains largely unchanged, as all changes are contained in the transfer functions themselves. The transfer $H_r(\mathbf{X_l}, \mathbf{X_m})$ from loudspeaker to microphone is replaced with the transfer from loudspeaker to loudspeaker $H'_r(\mathbf{X_l}, \mathbf{X_l})$. This is a special case as the distance between source and measurement is zero. This would make the acoustic transfer function go to infinity, as the calculation includes a division by the distance [20]. To prevent this, a modified ATF is taken where the direct path is not included. This modified ATF is

$$H'_r(\omega, \mathbf{X_l}, \mathbf{X_l}) = \sum_{i=1}^{\infty} \frac{\exp(j\frac{\omega}{c}u_i)}{4\pi u_i}\exp(-j\omega t). \tag{6.13}$$

The derivation of this modified ATF is given in Appendix A.

6.1.3. ADC

Converting the analogue signal into a digital one, the ADC is the last component in the signal chain before the DSP. Using the signal model from [18], the ADC output is given by

$$U_q[k] = U[k] + N_q[k], \tag{6.14}$$

Where $U_q[k]$ is the output, $U[k]$ is the input signal and $N_q[k]$ is the quantisation noise. This only holds under three assumptions. Firstly, the assumption is that the quantisation noise $N_q[k]$ is random and uncorrelated with the original signal $U[k]$. No clipping of the ADC is to occur as well, to keep the error bounded. Furthermore, the original signal $U[k]$ is assumed to be zero mean and stationary.

6.1.4. Complete signal model after the ADC

Combining (6.6), (6.12) and (6.14), the signal model after the ADC is:

$$\begin{aligned} U_q[k] = &\overbrace{H_a[k]S[k] + N_a[k]}^{S'[k]} \\ &+ H_m(k,\theta_s)\overbrace{\left(H_r'(k,\mathbf{X_l},\mathbf{X_l})(H_l[k](H_a[k]S[k] + N_a[k]) + N_l[k])\right)}^{S'''[k]} \\ &+ H_m(k,\theta_x)\overbrace{(H_r(k,\mathbf{X_x},\mathbf{X_l})X[k])}^{X'[k]} \\ &+ H_m(k,\theta_n)N_r[k] + N_m[k] + N_q[k]. \end{aligned} \tag{6.15}$$

6.2. Using a digital filter to remove unwanted components

A DSP algorithm can remove unwanted components from the incoming signal $u(t)$ to get an estimate of the desired signal. What the desired signal is can differ per use case, and four different desired outputs are discussed here. In all cases, the filter can be described with:

$$\hat{Y}[k] = H_f[k]U[k], \tag{6.16}$$

where $U[k]$ is the filter input as given in (6.15), $H_f(\omega)$ is the filter transfer function, and $\hat{Y}[k]$ is the estimate of the desired signal $Y[k]$.

Case 1: Estimation of a point source in the room at the loudspeaker location

In this case, the goal is to identify the point source in the room as given in (6.9) and recorded by the loudspeaker:

$$Y[k] = H_m(k,\theta_x)H_r(k,\mathbf{X_x},\mathbf{X_l})X[k]. \tag{6.17}$$

This means any form of the played signal has to be removed. Additional noise cancellation might need to be applied to obtain an accurate signal estimate for $\hat{Y}[k]$. If there is prior knowledge about the room acoustics, this could be integrated in the filter to remove both a playback signal on the line as well as the reflected playback signal.

Case 2: Estimation of the playback signal including room acoustics

Here, the goal of the intended output of the filter is to provide information on the playback of the signal. Any signal already present in the room is not of interest and the signal to be estimated is:

$$Y[k] = H_m(k,\theta_s)H_r'(k,\mathbf{X_l},\mathbf{X_l})(H_l[k](H_a[k]S[k] + N_a[k])). \tag{6.18}$$

Case 3: Estimation of the ambient noise level in the room

In some cases the ambient noise level in the room needs to be estimated, e.g. to be able to adjust incoming signals to be played. In this case the desired signal is

$$Y[k] = H_m(k,\theta_n)N_r \tag{6.19}$$

The recording device noise and quantisation noise are expected to remain in the filter output $\hat{Y}(k)$ as an unfortunate byproduct of the choice of $Y[k]$. This estimation can be done without any signal playing in the frequency band of interest. Prior knowledge of the room acoustics could also be used here.

Case 4: Estimation of the complete recording at the loudspeaker position

In this case the full recording is of interest, and only the direct playback signal needs to be removed. The desired signal looks like:

$$\begin{aligned} Y[k] =& H_m(k,\theta_s)H'_r(k,\mathbf{X_l},\mathbf{X_l})(H_l[k](H_a[k]S[k]+N_a[k])+N_l[k]) \\ &+ H_m(k,\theta_x)H_r(k,\mathbf{X_x},\mathbf{X_l})X[k] \\ &+ H_m(k,\theta_n)N_r[k] \end{aligned} \tag{6.20}$$

Here again, the noise from the different system components is still present in the filter output $\hat{Y}[k]$. Note that (6.17), (6.18) and (6.19) are all contained in (6.20). This means a filter that estimates $Y[k]$ as given in (6.20) is at the base of the other filters as well.

6.3. Summary

In this chapter a signal model framework for a playback and recording system using only one loudspeaker has been presented to answer research question **RQ 3a**. After this, different desired outputs for a digital filter have been identified and described. The signal model is used in a simulation for a proof-of-concept, described in Chapter 8.

7

Concept electronics system design

As the interface between the loudspeaker and the digital domain, the electronics used in the system are an essential part of this thesis. To properly acquire the recorded signal, an evaluation of the necessary electronics in the system is done. With this evaluation, this chapter provides an answer to research question 2c:

RQ 2c *- What electronic devices are needed for recording during playback?*

The theoretical information presented in this chapter can be used for the design of a physical system which can play and record audio using a single loudspeaker. A proper electronic model of a loudspeaker will also provide information that can be used in the signal model.

7.1. Electronic model of a moving coil loudspeaker

In Chapter 6 the loudspeaker model is presented as a block with a transfer function, along with several assumptions about the device. A more in-depth look at how a loudspeaker is modelled can improve the signal model by validating these assumptions.

7.1.1. Modelling the loudspeaker

As described in chapter 2 of [1], a loudspeaker motor can be modelled with an electronic equivalent circuit as shown in 7.1. In this model, v is the velocity of the cone, e_g is the voltage generated by the movement of the voice coil, and Z_{OUT} is the amplifier output impedance. As a conventional amplifier acts as a voltage source, Z_{OUT} this impedance disappears.

7.1.2. Information contained in the model

The model shown in Figure 7.1 consists only of R, L, C and M components, and as such the circuit is reciprocal [21]. This means the circuit is linear too, and the superposition principal holds for any sources added to the system. If we add the recording capabilities as a voltage source in the system, and we have the playback signal as a voltage source elsewhere, these sources can be evaluated independently from each other. From this the conclusion is drawn that during playback, the recording capabilities of the loudspeaker are not influenced by the signal from the amplifier and vice versa.

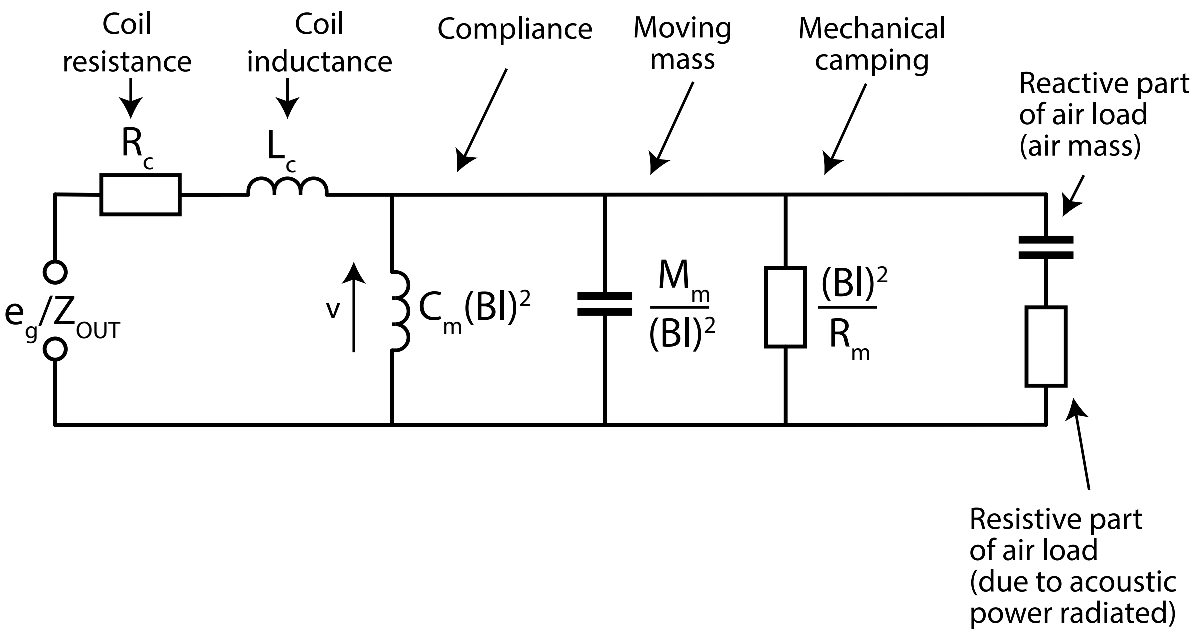
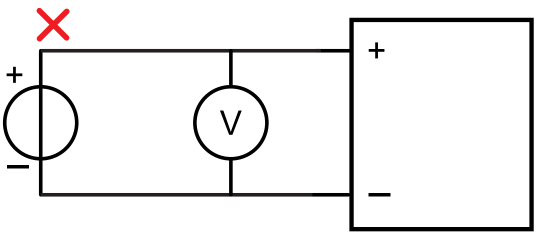


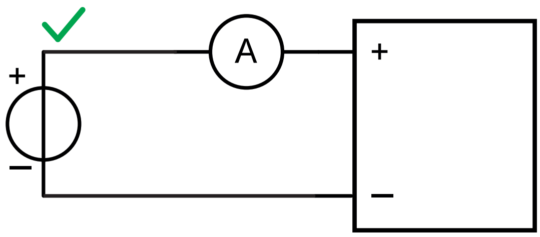
Figure 7.1: The equivalent circuit of a moving-coil drive unit which models both the operation of the transducer and the actual electrical load seen by the amplifier.

Determining the signal to be measured

Usually in audio applications, the signal voltage is the quantity that is being measured. However, as stated in Chapter 3, the loudspeaker is driven by a voltage amplifier. Measuring the voltage across the terminals of a loudspeaker then results in the situation shown in Figure 7.2a. It is evident that this will not yield a useful measurement, as the voltage source also directly drives the measurement. This means that the current needs to be measured, like shown in Figure 7.2b. This is an important design constraint on the measurement system so it needs to be taken into account.



(a) A representation of a voltage measurement on a voltage-driven device. This will result in only a measurement of the driving signal.



(b) A representation of a current measurement on a voltage-driven device. Any currents coming from the block on the right can be measured, together with the current induced by the driving voltage.

Figure 7.2: Schematic representation of a voltage measurement (a) and a current measurement (b) of a voltage-driven system. Only a current measurement will provide additional information about possible sources in the block on the right.

7.2. Concept circuit design: using a transimpedance amplifier

To provide insight into the design of an electronic system that is capable of recording and playing audio, some concept-level circuit design are presented here. These designs have been made to show the circuit topologies needed to achieve a successful recording without affecting the playback of a signal. These designs can be used to evaluate the alterations that need to be made to existing amplifier designs. Amplifier designs in PA systems are not publicly available, therefore no additional assessment can be performed at this point. However, it is possible to discuss the design considerations and topologies based on present knowledge from the Bosch PRAESENSA system.

As most ADCs measure a voltage across their inputs, the current needs to be converted into a voltage before it can be properly read. The simplest way to achieve this is with a shunt resistor. This, however, influences the circuit and will introduce some uncertainty in the measurement. Given the low recording output sensitivity of the loudspeaker and the high dynamic range of the signal, this method is not considered to be a suitable adaptation for the measurement.

Another way to get a voltage reading from a current is using a transimpedance amplifier. In this configuration, a current is converted to a voltage, with the added possibility of gaining the signal in the process. Such a configuration is shown in Figure 7.3. The impedance in the feedback path Z_f determines the gain of the amplifier, and this value can be chosen by the designer. To get a proper measurement over the audible frequency range, the impedance in the feedback path must follow the impedance of the loudspeaker over this range. As the loudspeaker impedance varies immensely over this range, matching the feedback impedance to this is not a trivial task. The first consideration could be to place another loudspeaker of the same model in the feedback path. In this way, the feedback impedance follows the same curve as the loudspeaker. This option is very impractical, as it offers no flexibility, and means adding a large component in the system.

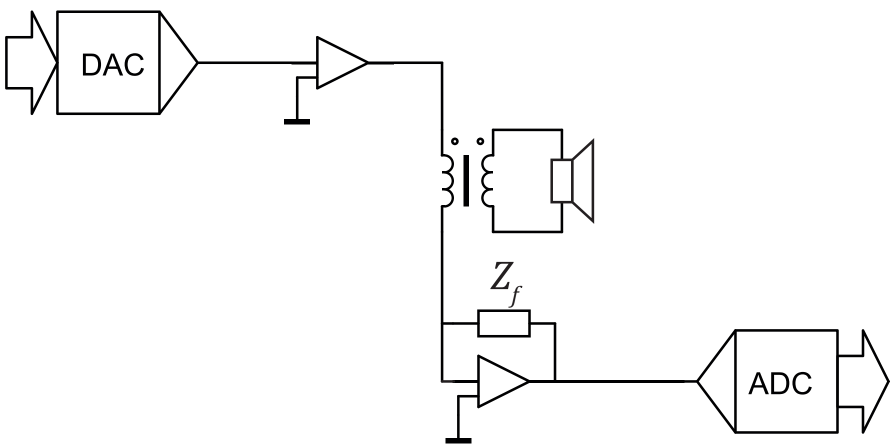


Figure 7.3: An electronic set-up utilising a transimpedance amplifier to allow for simultaneous playback and recording. Setting the feedback impedance Z_f is not a trivial task, however. The impedance curve of the loudspeaker will drastically influence the current going into the transimpedance amplifier.

Using a secondary amplifier channel could be a feasible solution for a demonstration set-up. This design is shown in Figure 7.4. The concept in this design is using a digital feedback system. The feedback impedance is set to get a static gain over the frequency range, and the frequency-dependent impedance of the loudspeaker is disregarded. A secondary output channel is then added together with the input of the transimpedance amplifier. The function of this secondary channel is purely to compensate for the impedance curve of the loudspeaker. A calibration algorithm can then be implemented in the digital domain. To calibrate, a frequency sweep is played over the output channel while no sound is being recorded. The secondary channel is then iteratively adjusted until the input at the ADC matches this sweep. This way the secondary channel will end up playing a signal to compensate the impedance of the loudspeaker over the frequency range.

If using a secondary channel is not an option, more involved analogue impedance compensation is necessary. This compensation is not discussed here, as the designs presented are of a concept level. More sophisticated electronics design is outside the scope of this thesis.

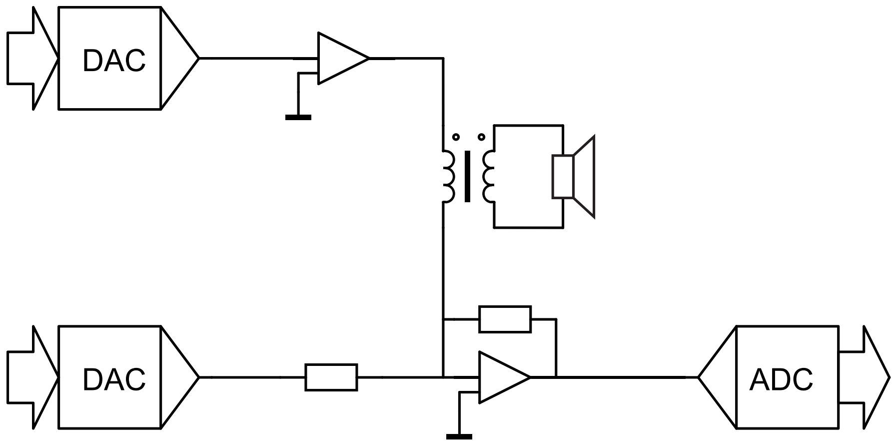


Figure 7.4: The use of a secondary amplifier channel will allow for the use of a digital feedback loop, which can be used to calibrate the system to remove the effect of the loudspeaker impedance on the recorded signal.

7.2.1. Choosing an ADC

As discussed in Section 5.2.3, the on-axis output sensitivity of the measured loudspeakers as microphone is between $0.39mV/Pa$ and $1.24mV/Pa$. The worst-case dynamic range was calculated to be $119.1dB$. To be able record this, the ADC needs to be capable of handling this range, or the recording is useless. ADCs exist that can handle this dynamic range [22]. However, it needs to be a carefully selected high-quality ADC as most can not handle this range.

7.3. Improving the dynamic range: using a differential amplifier

Another option is to decrease the dynamic range before the signal hits the ADC. For this, a circuit like in Figure 7.5 can be used. In this circuit the current and the voltage are both measured. As the voltage measurement is driven by the amplifier, this contains the signal played. The current measurement contains this as well, on top of the recording signal. Both of the measurements then go into a differential amplifier. Setting the first two amplifiers in the proper ratio, the playback signal will end up as the common-mode signal between the signals. A common-mode signal is a voltage that is equal (common) at both inputs of a differential amplifier. Designing the differential amplifier to have a high common-mode rejection ratio will then suppress the playback signal and lower the dynamic range of the input.

This design also has several drawbacks. As both the voltage and current are read, the impedance compensation as conceived in Section 7.2 will not work optimally with the voltage amplifier. It will skew the differential signal and the measurement will most likely be worse. This means the impedance compensation needs to be done in a different way. Another drawback is the increased component cost. Compared to the design from Section 7.2, this design includes more amplifiers. The trade-off is either carefully selecting a high-quality ADC, or designing more electronics but being able to use a possibly cheaper ADC with a lower input range. This trade-off should be made on a case-by-case basis.

The goal of this circuit is roughly the same as the digital filter described in Section 6.2. Equation (6.20) describes the filter output as the recorded signal without the playback signal, and this is what the circuit described in this section aims to give at the ADC input. This shows that analogue circuitry and signal processing go hand in hand, and there are often multiple ways to solve a problem. This thesis will continue with the implementation of a digital filter.

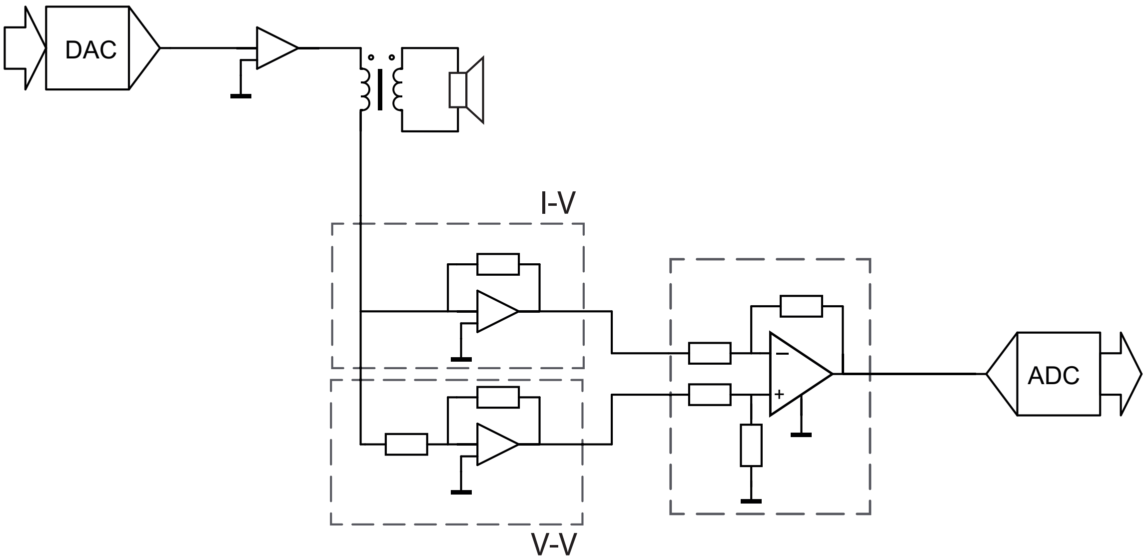


Figure 7.5: Using a transimpedance amplifier combined with a voltage amplifier, the playback signal could possibly be removed by making it the common-mode signal at the input of a third, differential amplifier. If this amplifier is designed to have a high common mode rejection ratio, the dynamic range of the recording could be improve before it hits the ADC. The impedance compensation does become a lot more involved by doing this.

7.4. Summary

In this chapter, the electronic equivalent model of a loudspeaker is analysed. From this, it is concluded that for the linear case, simultaneous recording and playback do not influence each other. This validates the assumption made in Chapter 6 on this. The signal recording possibilities are also discussed, and different concept-level electronic designs are presented. With this, research question **2c** is answered. To acquire the recorded signal from a voltage-driven system, the current needs to be measured. To achieve this, two possible designs with a transimpedance are presented, as well as one design utilising a differential amplifier design.

8

Use case simulation and filtering

This chapter discusses the attempted signal quality improvement through the use of digital signal processing (DSP). This answers research question 3b:

RQ 3b *- What algorithms can be used to improve the quality?*

To answer this research question, a simulation is set up in MATLAB to allow different use cases to be tested efficiently. A basic Least-Mean-Square (LMS) filter is then implemented to remove unwanted signal components after generating the signals from the simulation. To evaluate the results, a quality measure is defined.

8.1. Implementing a simulation based on the signal model

To set up a simulation, a simplified version of the signal model from (6.15) is taken. In (8.1) the simplifications are shown as crossed out with a horizontal line, after which the simplified version is shown in (8.2). The simplifications will be discussed here.

First, the amplifier is assumed to be ideal. This means $H_a = C$ and $N_a = 0$, so the signal is amplified by a frequency independent constant C with no noise. As the dynamic range is an important characteristic for the simulation, this is included in this constant. To simplify the amplifier transfer further, the constant is moved to the microphone transfer by multiplying it with $1/C$. In this way the amplifier transfer can be set as $H_a = 1$. The other devices are simplified to also not introduce noise, which means $N_l = N_m = N_q = 0$. Lastly, a simplification is done on the recording response. The angular responses in the frequency range of interest are relatively similar to each other, and signal reflections spread out the signals over a wide range of angles. To average this effect, the recording response can reasonably be simplified to just the on-axis response in all directions.

$$U = \cancel{H_a}S + \cancel{N_a} + H_m\cancel{(\theta_s)}\big(H_r'(\mathbf{X_l},\mathbf{X_l})(H_l(\cancel{H_a}S + \cancel{N_a}) + \cancel{N_l})\big) + H_m\cancel{(\theta_x)}(H_r(\mathbf{X_x},\mathbf{X_l})X) + H_m\cancel{(\theta_n)}N_r + \cancel{N_m} + \cancel{N_q} \tag{8.1}$$

$$U_s = \overbrace{S}^{\text{On the line}} + \overbrace{H_m H_r'(\mathbf{X_l},\mathbf{X_l})H_l S + H_m H_r(\mathbf{X_x},\mathbf{X_l})X + H_m N_r}^{\text{In the room}} \tag{8.2}$$

8.1.1. Choosing the signals to be used in the simulation

To represent the different use cases, the proper signals need to be used in the simulation. This section describes what signals were chosen and for what reason.

- $s(t)$: The announcement signal. These signals are mostly speech signals. Speech excerpts from the TIMIT corpus [23] were used as the announcement signal s. To be able to efficiently separate this from the point source signal $x(t)$ while listening, speech excerpts of a female voice are used.
- $x(t)$: A point source in the room. In the use case of event detection, one could be looking for people using the PA system in a building. Using a male speech sample from the TIMIT corpus, it will be easy to distinguish the point source signal from the female announcement signal while listening. Using a speech sample will provide some additional benefits. Speech enhancement algorithms are very common as a lot of research in signal processing focuses on this topic. Additional speech enhancement is thus possible if needed. Along the same line, speech intelligibility measures might prove to be useful in quality assessment. Lastly, using a speech signal here will also help exposing the potential privacy risk from espionage using PA systems.
- $n(t)$: Noise present in the room. To be able to simulate different use cases, different types of noise are chosen for N. In the performed simulations, an artificial non-stationary noise signal was used.

8.1.2. Using the playback and recording responses

For the loudspeaker recording and playback responses, the magnitude response data from Chapter 5 was used together with the filter design tool in MATLAB to generate a linear phase direct-form finite impulse response (FIR) filter. For the simulation presented here, the on-axis response of the LC1-WM06E8 #1 speaker were used.

As the recording response measurement started at 100Hz, where the response is still quite high, a high-pass filter was made to reduce the filter response at 100Hz to a low enough level (-80dB) to avoid ringing in the time domain. As the TIMIT speech files are recorded at a sampling frequency of 16kHz, any audio content above 8kHz can introduce aliasing. To remedy this, a low-pass filter was also designed to reduce the response significantly at 8kHz (-80db), so the recording response does not fold over on itself. The recording response was cascaded with the high-pass and low-pass filters to generate a final version. This is shown in Figure 8.1.

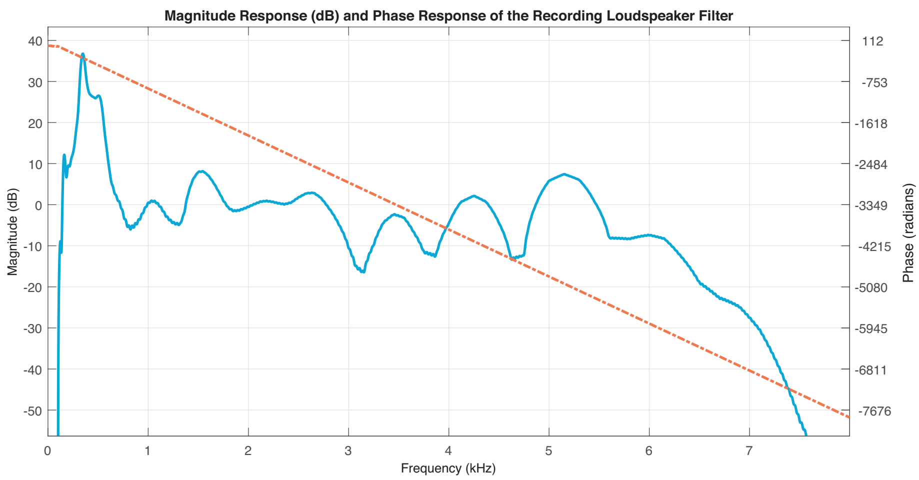


Figure 8.1: The magnitude response and phase response of the filter generated from the measurement data. This filter models a recording loudspeaker. Around 100Hz a high-pass filter is introduced to mask the start of the measurement data. Close to 8kHz a low-pass filter is introduced to prevent aliasing.

An adjustment similar to the one described above was also performed on the loudspeaker playback response. This response was defined down to 20Hz, where the transfer is low enough to not need a high-pass filter as with the recording response. The playback response was cascaded with the same low-pass filter as the recording response, to ensure no aliasing from content above 8kHz. This is shown in Figure 8.2.

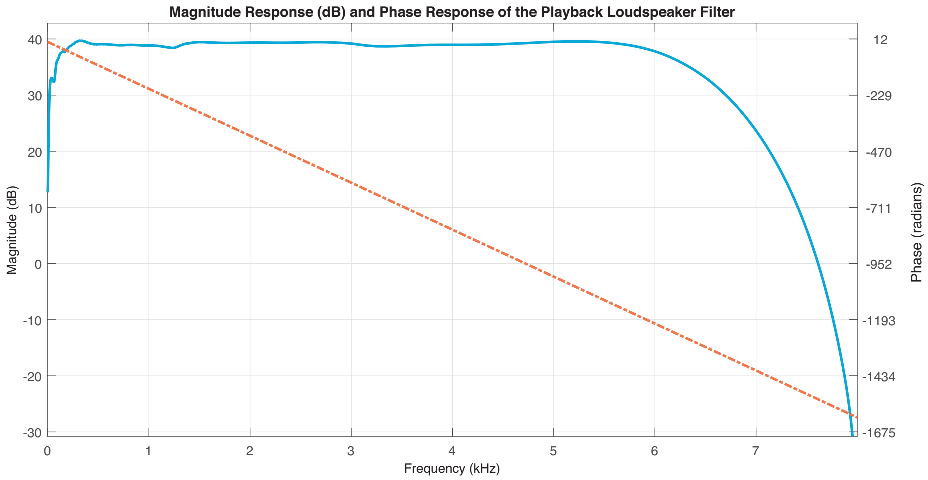


Figure 8.2: The magnitude response and phase response of the filter generated from the datasheet data. This filter models a loudspeaker used for playback. Close to 8kHz a low-pass filter is introduced to prevent aliasing.

8.1.3. Generating an acoustic transfer function

To describe a room through which the signals propagate, acoustic transfer functions (ATFs) are used. In [20], a method to efficiently simulate small-room acoustics using an ATF was introduced. To provide flexibility, the ATFs are generated using a modified version of the MATLAB mex function described in [24]. This function is based on the method presented in [20], and the modified version checks for a result of infinity, and removes those from the summation. Equation (6.13) describes the generation of the ATF, where it removes elements with zero distance. As the loudspeaker is simulated on the ceiling plane, more elements than only the direct path will result in infinity. Therefore these result also needs to be removed, and the modified code handles this accordingly.

As the LC1 range is a ceiling loudspeaker range, it is placed inside of the ceilings in office buildings, and specifically office rooms. To simulate an average office room, the rough dimensions from a faculty office were used: the simulated room dimensions are 5 m long by 4 m wide by 3 m high. In this room the ceiling loudspeaker is placed at position $(4, 2, 3)$: 1 m away from the wall where a door would be, centred in the room width and on the ceiling. The point source is modelled at $(3, 2, 1.8)$: an average height for a Dutch man, standing in the room. Finally, a noise source is also simulated as a point source. Generally noise sources are not point sources, but to reduce complexity the noise was placed at $(0, 1, 1)$: imagine noise from outside coming in from a small window. In (8.2), this results in a substition of N_r with N_r', where $N_r' = H_r(\mathbf{X_n}, \mathbf{X_l})N_r$

Using the generator code the ATFs were generated for a room with a relatively high reverberation. The reflection coefficients are $[\beta_{x1}, \beta_{x2}, \beta_{y1}, \beta_{y2}, \beta_{z1}, \beta_{z2}] = [0.5, 0.4, 0.45, -0.5, -0.4, -0.45]$.

8.1.4. Adding everything together

To be able to use the correct ratios of the signals, all the source signals were normalised first. For N and X this was straightforward, but the announcement signal S is filtered first with the filter modelling the loudspeaker transfer. After this the signals are scaled to their respective dB levels. For the loudspeaker

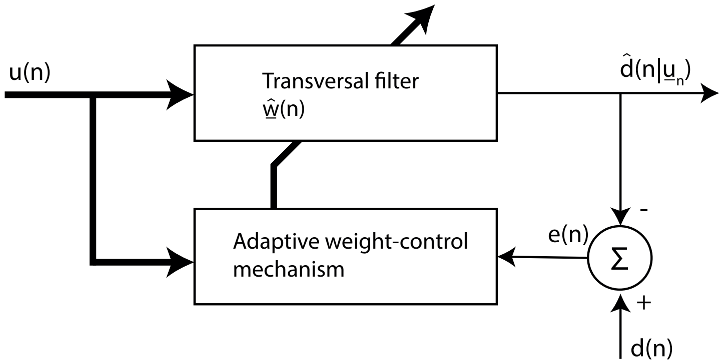


Figure 8.3: A block diagram showing the working of a Least-Mean-Square (LMS) filter.

signal S, 88 dB was used, as this is specified to be the SPL of the loudspeaker playing at a power of 1 W. This would be sufficiently loud for a space of the chosen dimensions. The point source signal X was scaled to 60 dB, the level of a regular conversation at 1 m. Finally the noise signal was set to be also 60 dB, the sound level of a fast car at roughly 50 m.

Following this scaling, the signals are convolved with their respective ATFs and the microphone response, as to generate the in-room part of Equation 8.2:

$$H_m H_r'(\mathbf{X_l}, \mathbf{X_l}) H_l S + H_m H_r(\mathbf{X_x}, \mathbf{X_l}) X + H_m N_r \tag{8.3}$$

To be able to add the signal S as it is on the loudspeaker line, the recorded signal is normalised again. This allows us to add the two signals together in the specified dynamic range. For the loudspeaker playing at 1 W, the dynamic range is calculated to be 105.6 dB. The calculation is the same as in Section 5.2.3 (where it was 116.7 dB), but with a power of 1 W instead of 6 W. As a final step the signal is normalised again, to ensure no clipping occurs when listening to the signal or when writing it to an audio file.

8.2. Estimation based on the recorded signal: a proof-of-concept

To evaluate whether using a digital filter on the recorded signal is a feasible operation, a proof-of-concept needs to be performed. To accomplish this, a simple Least-Mean-Square (LMS) filter is implemented. This section describes the use of this filter, as well as the software implementation.

8.2.1. Least-Mean-Square filter implementation

The LMS filter [25] is an adaptive filter, based on a series of delayed weights. It optimises the filter weight by minimising the error between the filter output and the desired filter output. It is based on a transversal filter structure, where the weights for each delay are controlled by the algorithm. The block diagram in Figure 8.3 shows an illustration of this. The iterative update process from the LMS filter can be described as follows [26]:

1. Filter output:
$$\hat{y}(m) = \hat{\mathbf{w}}^H(m) u(m), \tag{8.4}$$
where m is the filter tap index up to order M, $\hat{\mathbf{w}}(m)$ are the filter weights (usually initialised at zero), and $u(m)$ is the filter input.

2. Estimation error:
$$e(m) = d(m) - y(m), \tag{8.5}$$
where $d(m)$ is the desired filter output

3. Tap-weight adaptation:

$$\hat{\mathbf{w}}(m+1) = \hat{\mathbf{w}}(m) + \mu \mathbf{u}(m) e^*(m), \tag{8.6}$$

where μ is the update step size. Because in this simulation the input- and output signals are real-valued, $e^*(m) = e(m)$.

Usually, when using an adaptive filter like this, some prior knowledge of the desired signal is present. In this case though, the only signals known are the recorded signal and the playback signal. Still the adaptive filter can be used in a creative way. If the recorded signal ($u_s(n)$ with time index n) is used as the filter input, and the playback signal $s(n)$ is used as the desired signal, the filter will try to remove any other components from the signal. As the filter output tries to converge to the desired signal $s(n)$, the error signal will contain the recorded signal. Taking this error signal as the output of the filter will give the original signal with the playback signal adaptively removed from it.

8.3. Defining a quality measure

Different desired output signals have different characteristics. As quality is an objective term, the quality of the different signals could be based on various attributes. In this section, a quality measure is defined to help evaluate the performance of the simulated system.

The first quality measure is based on intelligibility. In the case where the goal is to estimate a speech source in a room, intelligibility could be a suitable measure to determine the quality. The amount of intelligibility gives a good representation of the amount of information that can be extracted from this signal. The quality measures can be described by the following questions:

- How much does the intelligibility of the signal increase by using a digital filter to remove unwanted components?
- How much is the intelligibility degraded by using a loudspeaker for recording compared to a microphone?

As an intelligibility metric, Speech Intelligibility In Bits (SIIB) is used [27]. Compared to other intelligibility metrics, SIIB has reasonably high correlation to real intelligibility [28]. As the data from this simulation differs from the data sets used for evaluation of SIIB, the output might not be entirely accurate. It is, however, sufficient for the comparison needed to answer the questions above. Care should be taken if one wants to compare the results to different data sets or other research.

SIIB provides results as the information rate in bits/second (b/s), based on the mutual information between a message M_t and a received signal Y_t. Generally, the information rate will be between 0b/s (zero intelligibility) and 150b/s (high intelligibility). As the intelligibility approaches 100%, the SIIB results become asymptotic. Any rate above 150b/s can thus be considered as close to 100% intelligibility.

8.4. Results and evaluation

For evaluating the improvement of the speech signal, the intelligibility of the male speech was calculated for three different signals. The first was the full signal U_s, as recorded. The second signal is the filtered signal, where the signal component on the line is removed. This simulates Case 4 in Section 6.2. The third signal is a more involved filter. If there is some information available about the room acoustics, this could then be used to improve the filter. This simulates Case 1 in Section 6.2. To simulate this case, the simulated reflections of the announcement signal are added to the LMS filter input:

$$d(n) = s(n) + s'''(n). \tag{8.7}$$

As the SIIB algorithm compares the output signal to a reference message, this reference also needs to be chosen. In the evaluation, the filter output is compared to the speech signal as it arrives at the loudspeaker. The reference signal $M(n)$ is

$$M(n) = h_r(\mathbf{X_x}, \mathbf{X_l}) * x(n). \tag{8.8}$$

The values are shown in Table 8.1. As can be seen, the filter almost always improves the intelligibility of the signal. If the filter only removes the announcement present on the line, the announcement still plays in the room,. This signal is louder than the speech signal source, so the speech signal is masked by the announcement signal. As a result, the intelligibility improves little or not at all. For both cases, the improvement is noticeable when also attempting to remove the reflections from the announcement. What also can be seen from these results is that there is quite a significant drop in intelligibility when switching from a microphone to a loudspeaker. This result is expected, as the recording response is severely different. The SIIB measure includes a model of the human auditory system [27], and the human hearing is less sensitive to lower frequencies, where the loudspeaker is most sensitive. This degrades the intelligibility severely.

Signal type	Intelligibility of speaker with microphone reference (in bits)	Intelligibility of microphone with microphone reference (in bits)
Full recording	5.93	5.90
Signal on line removed	5.66	17.87
Signal + reflections removed	7.98	341.75

Table 8.1: The speech intelligibility in bits (SIIB) for different speaker recording signals, as well as a microphone signal. Compared to the ideal recording from a microphone, a speaker performs significantly worse. In both cases however, the filter increases the intelligibility.

8.5. Conclusion

This chapter has described, analysed and implemented algorithms to attempt to answer research question 3b:

RQ 3b *- What algorithms can be used to improve the quality?*

The simulation has shown a basic LMS filter can be used to improve the signal quality of the recorded signal by removing unwanted components. Also it has shown that prior knowledge such as having the information on the ATF $H_r'(\mathbf{X_l}, \mathbf{X_l})$ can be used to improve the filter performance significantly. As the simulation uses only a basic filter, it is reasonable to assume more advanced signal processing techniques could improve the result even more. Therefore, a successful proof-of-concept is achieved.

In this thesis, the use of more involved techniques is left as a recommendation for future work. Some techniques that might be helpful are acoustic echo cancellation [29],or the generalised cross-correlation phase transform (GCC-PHAT) algorithm [30]. In addition to this, it could be possible for the DSP to compensate for the loudspeaker recording transfer, so the signal quality is further improved. For the case of a basic LMS filter the speech output still has noticeable artefacts, and compensation is deemed not feasible, as it would also amplify these artefacts. Another interesting topic is the estimation of the ATF. It is expected this can be done using the system without any extra components. However, this does imply compensation for the loudspeaker recording transfer is in place. After this it should be similar to any other ATF measurement.

9 Conclusion and recommendations

The goal of this thesis was proving the feasibility of using a loudspeaker as a recording device in public address (PA) systems. The work presented covers an overview on a system which can simultaneously play and record audio over a loudspeaker, where a test set-up was used to most accurately simulate the system as those found in typical PA systems.

To answer the research question in a practical setting, two use cases were defined:

- *Event detection:* The system aims to estimate the presence and position of people in emergency situation, or to estimate a time of occurrence for a certain event such as an explosion.
- *Adaptive announcements:* The system aims to improve the signal quality and intelligibility of announcements by estimating the room acoustics and the ambient noise level.

The information found by answering the research questions provided a sufficient basis to prove the feasibility of recording with a loudspeaker in PA systems for these two use cases. Using a current measuring set-up with an analogue-to-digital converter capable of detecting a range of roughly 120 dB, a speech signal up to three metres in a cone of 120° from a loudspeaker can be successfully estimated in an office room with an announcement playing and background noise present. As the estimated signal is unknown to the system, the solution generalises to other signal types as well. A system with a single loudspeaker can be utilised for the use cases presented and is therefore proven feasible.

Recommendations for future research are largely based in two areas to improve the practicality of the system. The first is the signal quality improvement such that the range can be extended, and the system becomes applicable in a wider range of environments.and The second is focused on the additional issues arising from the implementation into a full PA system. As described in chapter 4, the isolated case of a single loudspeaker and amplifier system is considered. A full PA system would require the signals from different loudspeakers on a line to be isolated in signal processing, as well as communication of the data from the amplifier to the system controller for logging.

Recommendations for further research into the first improvement area, increasing the signal quality are:

- Improving the construction of a loudspeaker to have a more flat recording response without significantly altering the playback response,
- Designing a differential amplifier based on the concept shown in Section 7.3 to improve the dynamic range of the signal before going to the analogue-to-digital converter.

- Improving the signal processing with more involved digital filters on the signals recorded by loudspeakers will likely provide improved signal quality as the proof-of-concept has already shown promising results.
- In the use case of ambient noise level estimation, one could also research applying noise-tracking techniques such as presented in [31] to a filtered signal to improve the quality of the estimate further.

Recommendations for further research into the second improvement area, practical implementation of the system into a full PA system are:

- Research the generalisation of the Bosch PRAESENSA system to other systems. Other systems will have deviations in their implementations of the system and the recording system might need adjustments to accommodate for this.
- Evaluate and test the topology level designs shown in Chapter 7 for feasibility.
- Evaluate compensation methods for the loudspeaker recording transfer. Equalisation based on estimates of the loudspeaker recording transfer can be evaluated for implementation over a wide range of devices.
- Extending the single loudspeaker case from this thesis to the case where there are multiple loudspeakers connected to a line. Blind source separation techniques are expected to fulfil a pivotal role in this topic.

As the system of using a loudspeaker as a recording device has proven to be feasible, one important additional use case in Chapter 2, around espionage can also be revisited. Given the sensitivity of this topic, it is reasonable to assume that most research on this use case has been performed outside of the public domain. However, with the knowledge of the findings of this thesis, it can be safely assumed that an actor could use the key concept of this thesis for espionage.. The implications of this can best be illustrated by considering the requirements of performing espionage by using loudspeakers as microphones. There are three primary scenarios that can be considered here:

1. Actor has access to the PA system: To record a signal from the loudspeaker in a PA system that is not connected to the internet, direct physical access to the system is needed. Line monitoring systems present in the PA systems will detect changes in impedance or continuity in the line, and therefore the best access point is digitally at the amplifier or system controller to access a recording. Considering the level of access and knowledge about the system required, it becomes more likely that the actor would be setting up a microphone. Therefore this method of espionage, while possible, is likely not the largest security risk.
2. A potential larger risk is having direct access to signals of the system via the internet. This would allow an actor to remotely access the systems and record the signals. There are several practical implications to this method such as the required software or firmware of the software to be able to record, however once access to the system is possible these are challenges that can be overcome. Therefore this method poses a security risk. This also came up in conversation with Bosch Security Systems, where they explicitly stated that they employ extra, specifically designed, security measures to ensure no actors are able to access their systems over the internet. If a system with lesser protection is used, this could pose a serious threat.
3. Without access to the system, an actor would have to resort to optical or ultrasound methods to try and read the excursion of the loudspeaker. While this method is used in espionage already by measuring the excursion of for example a window or even a light bulb [32], there are other practical considerations that make this method of espionage improbable. Firstly, most PA systems have a grill in front of the loudspeakers making it hard or impossible to measure excursion. Secondly bringing a device into the space that can measure the excursion of a loudspeaker at distance is likely more complex than bringing an actual recording device into or near the space. With these

factors combined, it becomes highly unlikely that this method poses a greater security risk than already having a light bulb in the room.

All the cases described have severe drawbacks that limit their feasibility in this use case, therefore the security risk from recording audio over PA system loudspeakers is deemed to be low. Nevertheless, as part of the goal of espionage is to perform it unnoticed, it would be advisable to consider the specific circumstances of the cases described above and act accordingly to minimise potential security risks.

A

Special case ATF derivation

The special case ATF (6.13) is based on the mirror image source method (MISM) introduced by Allen and Berkley in 1979 [20]. In a rectangular room with rigid walls the sound pressure at the microphone is modelled as:

$$P(\omega, \mathbf{X}, \mathbf{X'}) = \sum_{p=1}^{8} \sum_{\mathbf{r}=-\infty}^{\infty} \frac{\exp(j\frac{\omega}{c}|\mathbf{R_p} + \mathbf{R_r}|)}{4\pi|\mathbf{R_p} + \mathbf{R_r}|} \exp(-j\omega t) \tag{A.1}$$

where c is the speed of sound in the room, $\mathbf{R_p}$ represents the vectors given by the eight permutations over $\pm$ of

$$\mathbf{R_p} = (x \pm x', y \pm y', z \pm z'), \tag{A.2}$$

$\mathbf{r}$ is the integer vector triplet (n, l, m), and

$$\mathbf{R_r} = 2(nL_x, lL_y, mL_z), \tag{A.3}$$

with (L_x, L_y, L_z) being the room dimensions. **X** and **X'** are the source and microphone position vectors (x, y, z) and (x', y', z'), respectively.

In the special case from section 6.1.2, the distance between the source and the microphone is zero. This will result in an occasional zero in the denominator for the direct path when $\mathbf{r} = \mathbf{0}$, and the impulse response becomes infinity in the limit. To suppress this effect, the direct path needs to be removed from the equation. We can rewrite (A.1) to have one summation:

$$P(\omega, \mathbf{X}, \mathbf{X'}) = \sum_{i=0}^{\infty} \frac{\exp(j\frac{\omega}{c}u_i)}{4\pi u_i} \exp(-j\omega t) \tag{A.4}$$

where **u** is a vector that contains all values $|\mathbf{R_p} + \mathbf{R_r}|$ in increasing order of magnitude. The direct path is contained in $u_0 = 0$, so we would like to remove the first element in the summation.

$$P'(\omega, \mathbf{X}, \mathbf{X'}) = \sum_{i=0}^{\infty} \frac{\exp(j\frac{\omega}{c}u_i)}{4\pi u_i} \exp(-j\omega t) - \frac{\exp(j\frac{\omega}{c}u_0)}{4\pi u_0} \exp(-j\omega t) \tag{A.5}$$

The summation can be rewritten to extract the $k = 0$ case:

$$P'(\omega, \mathbf{X}, \mathbf{X'}) = \sum_{i=1}^{\infty} \frac{\exp(j\frac{\omega}{c}u_i)}{4\pi u_i} \exp(-j\omega t) + \frac{\exp(j\frac{\omega}{c}u_0)}{4\pi u_0} \exp(-j\omega t) - \frac{\exp(j\frac{\omega}{c}u_0)}{4\pi u_0} \exp(-j\omega t) \tag{A.6}$$

This expression is not valid, as the last two terms still have zero in the denominator. Taking the limit of these terms gives:

$$P'(\omega, \mathbf{X}, \mathbf{X'}) = \sum_{i=1}^{\infty} \frac{\exp(j\frac{\omega}{c}u_i)}{4\pi\mathbf{u}_k}\exp(-j\omega t) + \exp(-j\omega t)\lim_{u_0\to 0}\left(\frac{\exp(j\frac{\omega}{c}u_0)}{4\pi u_0} - \frac{\exp(j\frac{\omega}{c}u_0)}{4\pi u_0}\right) \tag{A.7}$$

As the terms in the limit have the same convergence rate, the limit converges to zero. The resulting ATF for the special case is:

$$P'(\omega, \mathbf{X}, \mathbf{X'}) = \sum_{i=1}^{\infty} \frac{\exp(j\frac{\omega}{c}u_i)}{4\pi u_i}\exp(-j\omega t) \tag{A.8}$$

In some cases, e.g. if the device is on one of the axes, one will have more than one instance where $u_i = 0$. Because of the nature of **u**, these will be contained in the first elements. The method described above can be repeated until the summation starts at the first non-zero entry of **u**.

B

Loudspeaker measurement responses

B.1. On-axis microphone and loudspeaker responses

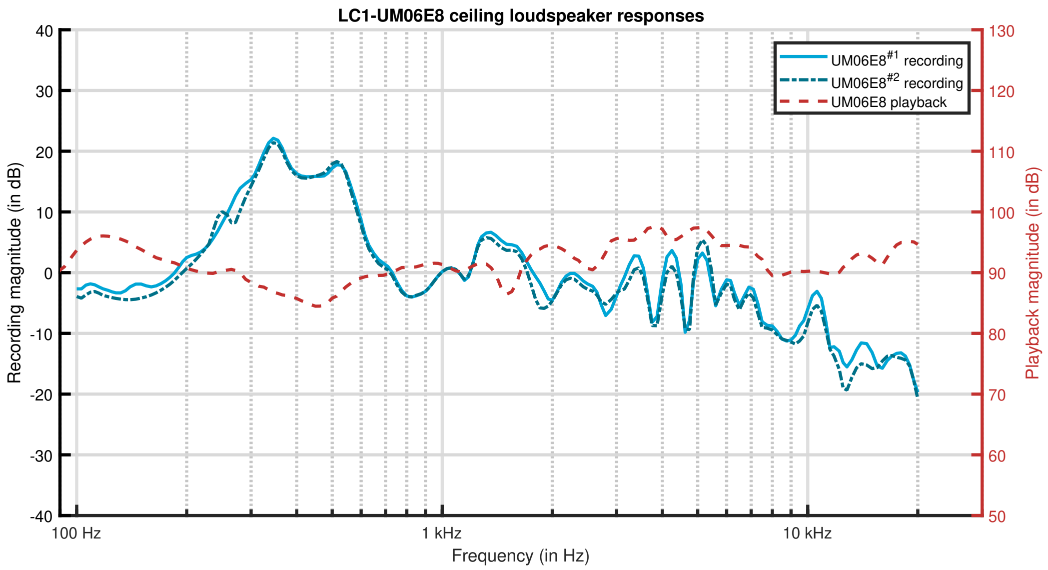


Figure B.1: The measured microphone response of the UM06E8 loudspeaker, compared with the loudspeaker response. The blue solid and dash-dotted curves represent the microphone response of two different loudspeakers. The red dashed curve is the loudspeaker response, as given by the datasheet.

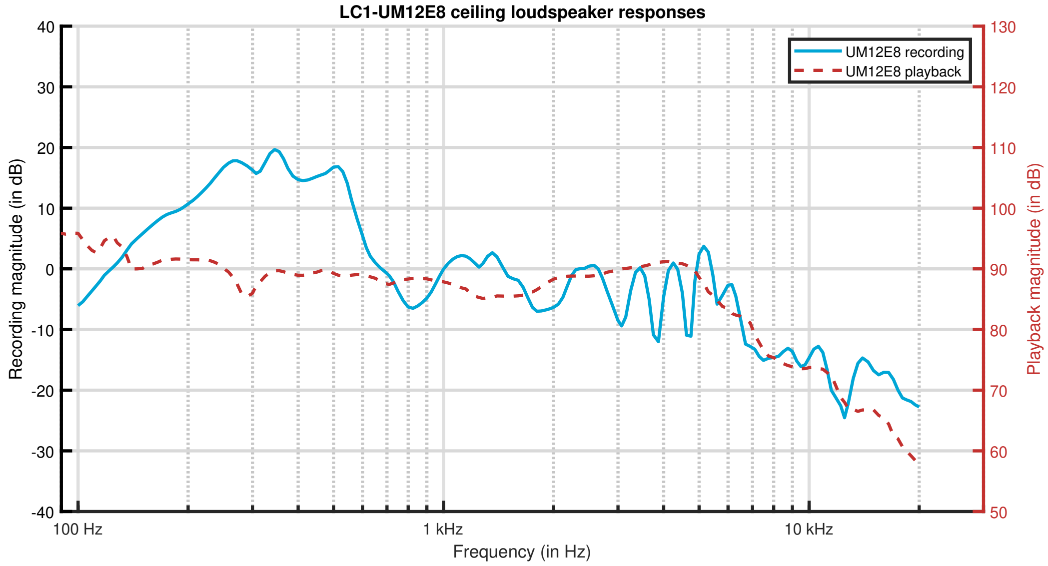


Figure B.2: The measured microphone response of the UM12E8 loudspeaker, compared with the loudspeaker response. The blue solid and dash-dotted curves represent the microphone response of two different loudspeakers. The red dashed curve is the loudspeaker response, as given by the datasheet.

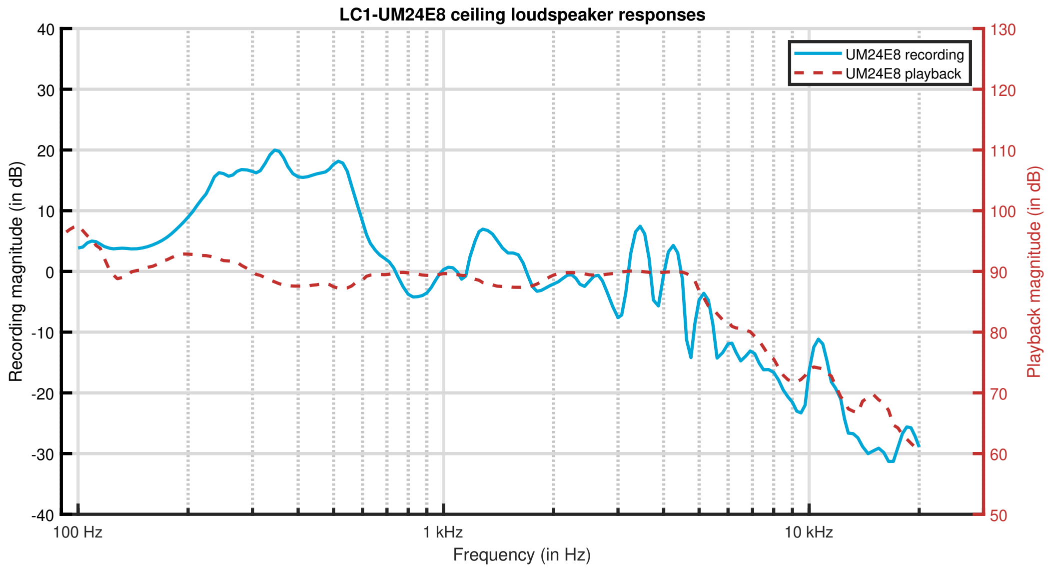


Figure B.3: The measured microphone response of the UM24E8 loudspeaker, compared with the loudspeaker response. The blue solid and dash-dotted curves represent the microphone response of two different loudspeakers. The red dashed curve is the loudspeaker response, as given by the datasheet.

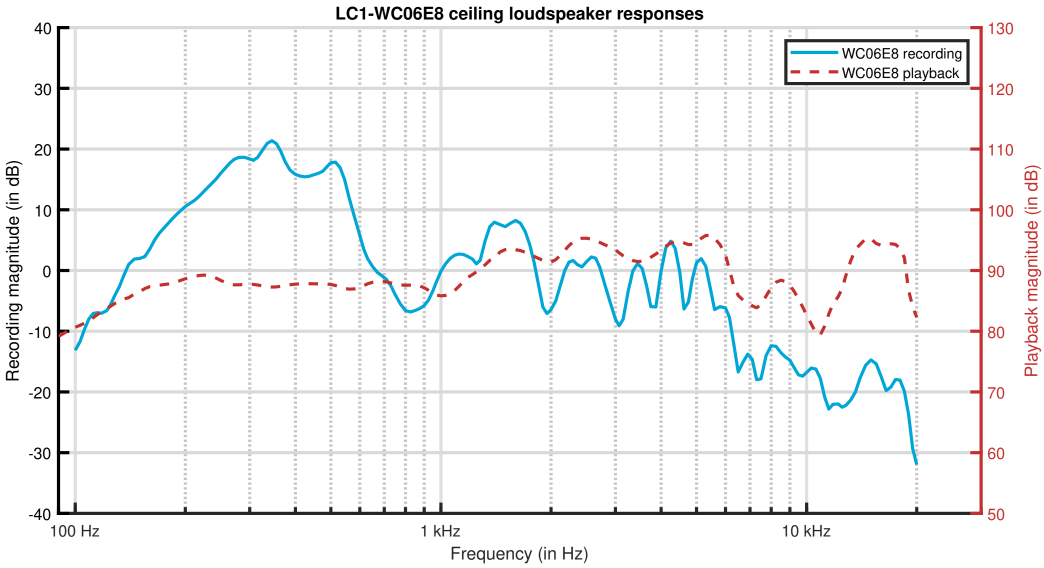


Figure B.4: The measured microphone response of the WC06E8 loudspeaker, compared with the loudspeaker response. The blue solid and dash-dotted curves represent the microphone response of two different loudspeakers. The red dashed curve is the loudspeaker response, as given by the datasheet.

B.2. Angular microphone responses

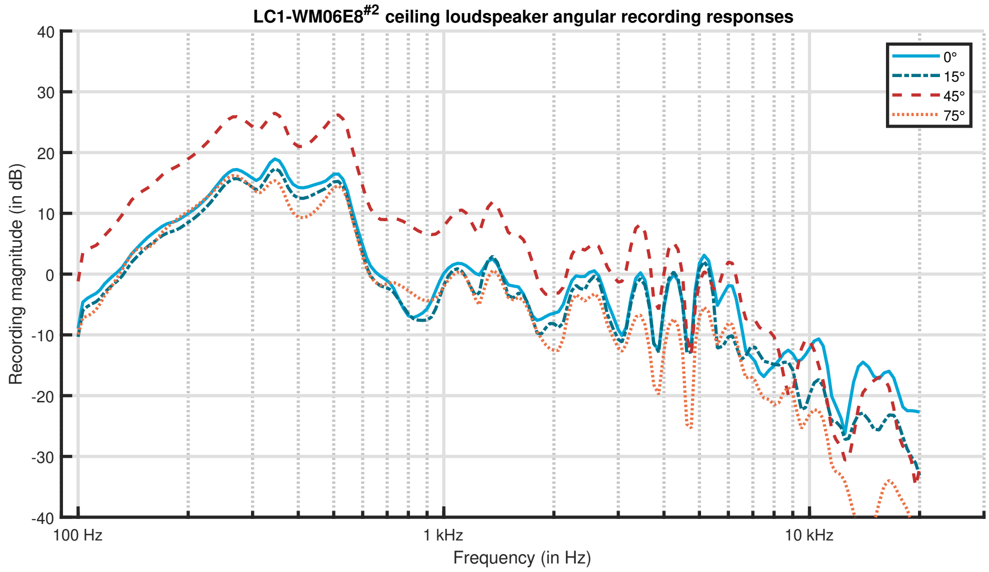


Figure B.5: The measured microphone response of the second WM06E8 loudspeaker over 4 different angles ($0°, 15°, 45°, 75°$).

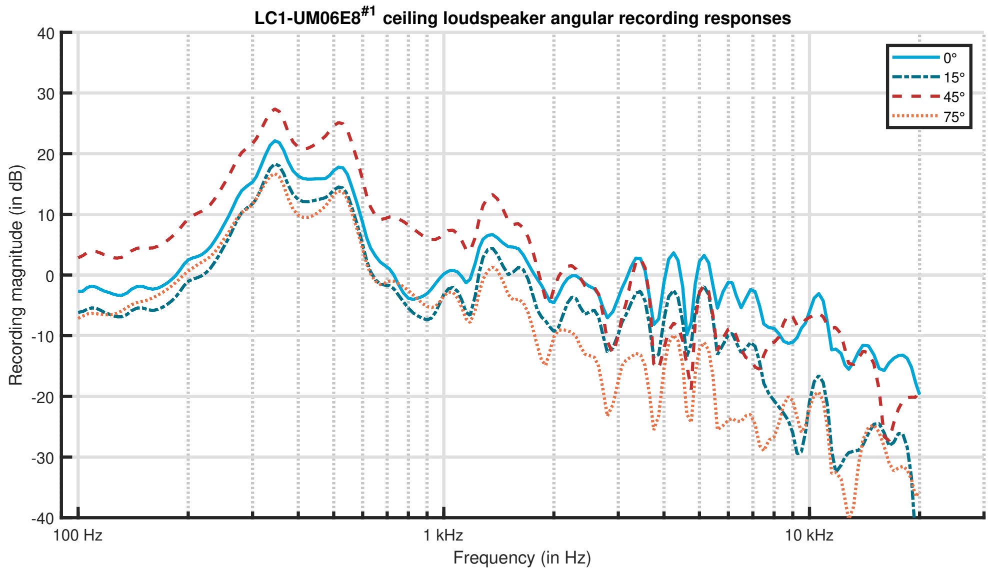


Figure B.6: The measured microphone response of the first UM06E8 loudspeaker over 4 different angles ($0°, 15°, 45°, 75°$).

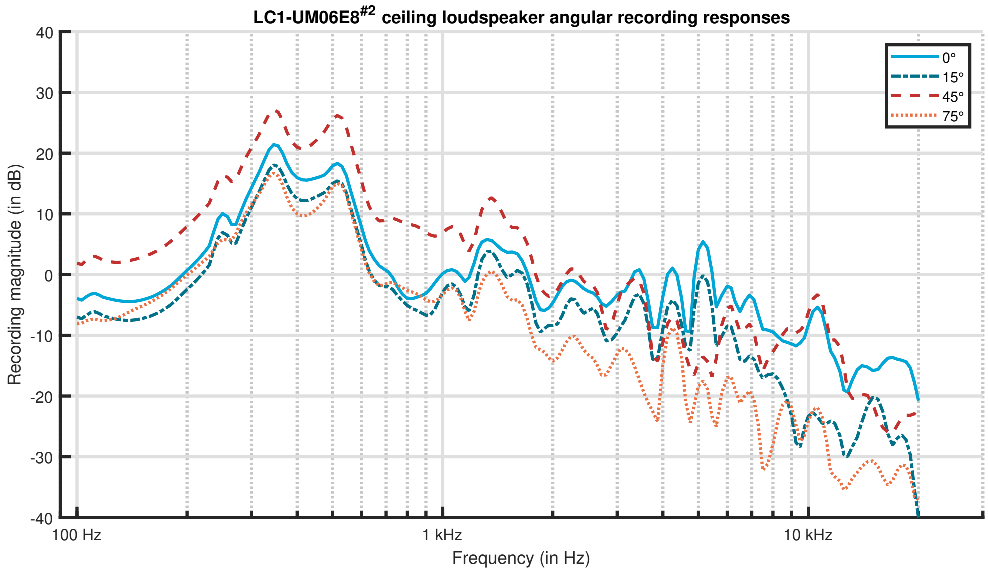


Figure B.7: The measured microphone response of the second UM06E8 loudspeaker over 4 different angles ($0°,15°,45°,75°$).

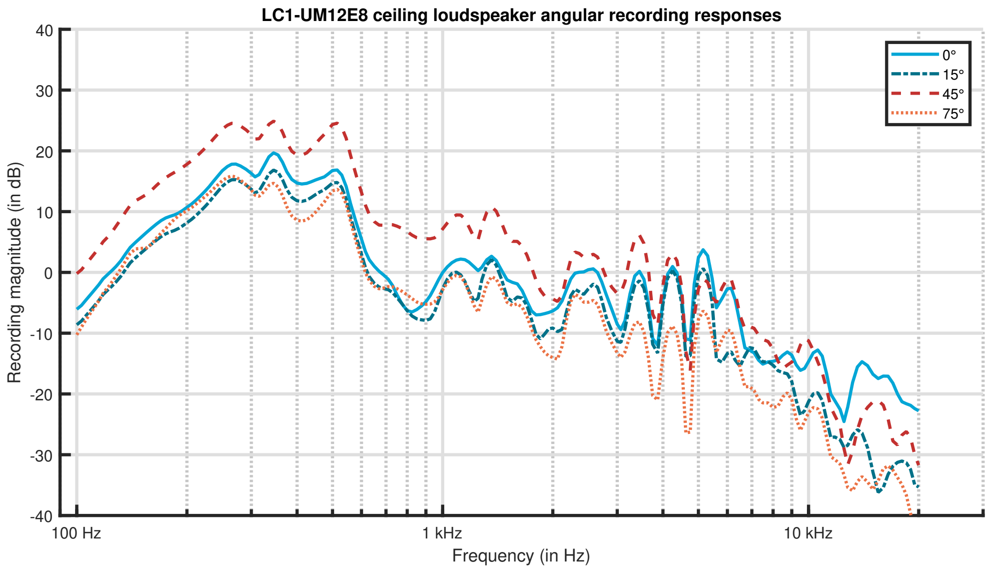


Figure B.8: The measured microphone response of the UM12E8 loudspeaker over 4 different angles ($0°,15°,45°,75°$).

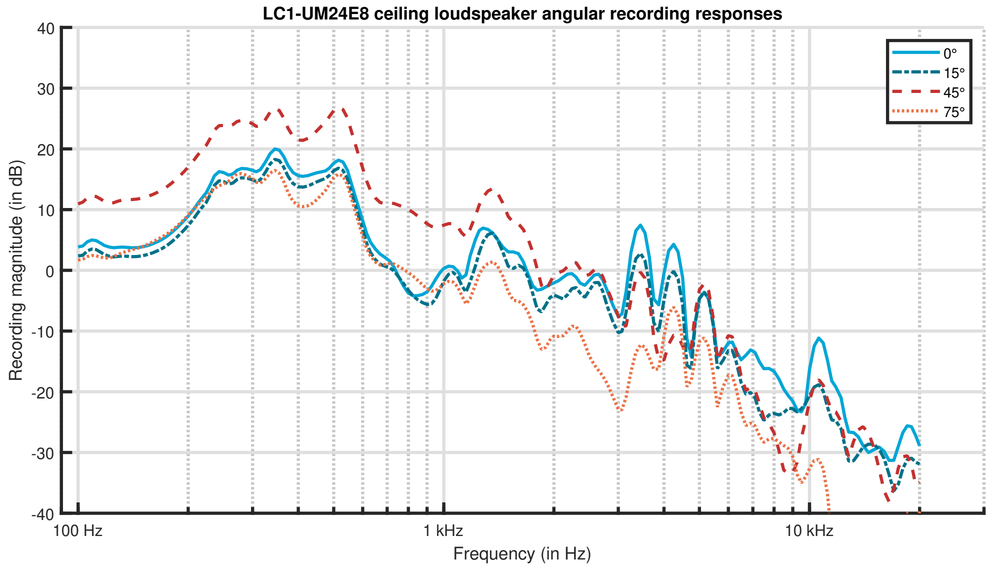


Figure B.9: The measured microphone response of the UM24E8 loudspeaker over 4 different angles $(0°, 15°, 45°, 75°)$.

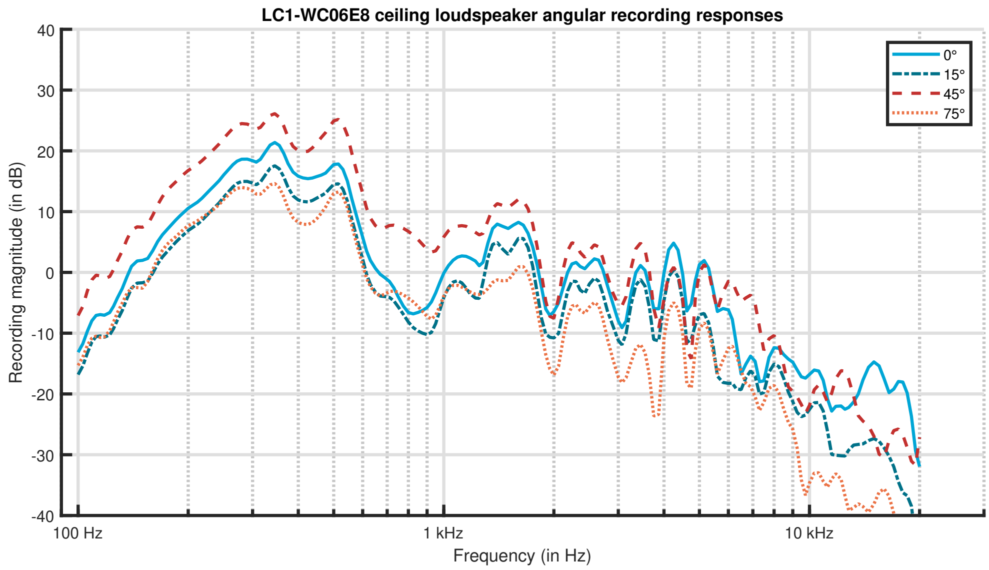


Figure B.10: The measured microphone response of the WC06E8 loudspeaker over 4 different angles $(0°, 15°, 45°, 75°)$.

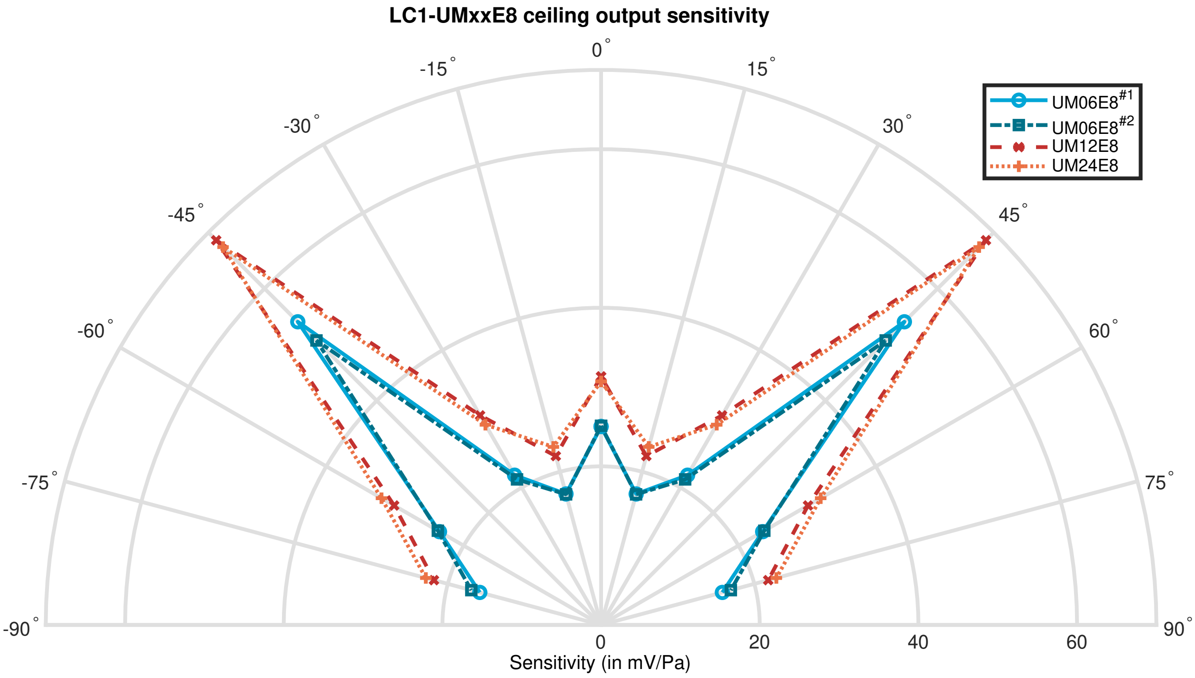


Figure B.11: Polar sensitivity plot of 3 loudspeakers (UM06E8, UM12E8 & UM24E8).

Bibliography

[1] J. Borwick, *Loudspeaker and headphone handbook*. Oxford: Focal Press, 2001.

[2] M. R. Gander, "Fifty years of loudspeaker developments as viewed through the perspective of the audio engineering society," *AES: Journal of the Audio Engineering Society*, vol. 46, pp. 43–50, Jan./Feb. 1998.

[3] J. Eargle and M. Gander, "Historical perspectives and technology overview of loudspeakers for sound reinforcement," *AES: Journal of the Audio Engineering Society*, vol. 52, pp. 412–432, Apr. 2004.

[4] J. Eargle, *The Microphone Book*. 2nd ed. Oxford: Focal Press, 2005.

[5] "IEEE Standard American National Standard Canadian Standard Graphic Symbols for Electrical and Electronics Diagrams (Including Reference Designation Letters)," *IEEE Std 315-1975*, 1993.

[6] H. Robjohns, "Can I make a 'Subkick' mic from any speaker cone." SoundOnSound.com. `https://www.soundonsound.com/sound-advice/q-can-i-make-subkick-mic-any-speaker-cone`. accessed May 17, 2021.

[7] E. Golden, "How To Use Headphones As A Mic." DJ techtools.com. `https://djtechtools.com/2015/02/22/how-to-use-headphones-as-a-mic/`. accessed Oct. 10, 2021.

[8] N. Le Goff, J. Jensen, M. S. Pedersen, and S. L. Callaway, "An introduction to OpenSound Navigator™," *Oticon A/S, white paper*, pp. 1–9, 2016.

[9] "Apple AirPods 3rd Generation." Apple.com. `https://www.apple.com/airpods-3rd-generation/specs`. accessed Nov. 22, 2021.

[10] Bosch Security Systems B.V., "PRAESENSA, IP-connected and fully-featured," 2018.

[11] Bosch Security Systems B.V., "PRAESENSA public address and voice alarm system," 2020.

[12] Bosch Security Systems B.V., "LC1 Modular Ceiling Loudspeaker Range," 2017.

[13] Bosch Security Systems B.V., "LBC 3018/01 Metal Cabinet Loudspeaker," 2016.

[14] R. M. Bews and M. J. Hawksford, "Application of the geometric theory of diffraction (GTD) to diffraction at the edges of loudspeaker baffles," *AES: Journal of the Audio Engineering Society*, vol. 34, pp. 771–779, Oct. 1986.

[15] J. Eargle, "The Pressure Gradient Microphone," in *The Microphone Book*, 2nd ed. Oxford: Focal Press, 2005, ch. 4.

[16] J. Stewart, "Performance of loudspeaker whizzers," *AES 110th convention*, May 2001.

[17] Kwan Chiu Radio MFG. Co., LTD., "LC1 transformer datasheet," 2011.

[18] J. G. Proakis and D. G. Manolakis, "Sampling and Reconstruction of Signals," in *Digital Signal Processing*, Prentice-Hall, 3rd ed., 1986.

[19] A. J. Montagne, *Structured Electronic Design*. 1.4 ed. Delft: Delft Academic Press, 2021.

[20] J. B. Allen and D. A. Berkley, "Image method for efficiently simulating small-room acoustics," *Journal of the Acoustical Society of America*, vol. 65, pp. 943–950, Apr. 1979.

[21] K. S. Suresh Kumar, “Two-Port Networks and Passive Filters,” in *Electric circuits and networks*, New Delhi: Pearson India, 2009.

[22] “Audio ADCs - Products.” `https://www.ti.com/audio-ic/converters/adc/products.htm`, 2021. accessed Dec. 1, 2021.

[23] J. Garofolo, L. Lamel, W. Fisher, J. Fiscus, D. Pallett, N. Dahlgren, and V. Zue, “TIMIT Acoustic-Phonetic Continuous Speech Corpus,” 1993.

[24] E. Habets, “Room impulse response generator,” *Internal Report*, pp. 1–21, Jan. 2006.

[25] B. Widrow, C. S. Williams, J. R. Glover, J. M. McCool, R. H. Hearn, J. R. Zeidler, J. Kaunitz, E. Dong, and R. C. Goodlin, “Adaptive Noise Cancelling: Principles and Applications,” *Proceedings of the IEEE*, vol. 63, no. 12, pp. 1692–1716, Dec. 1975.

[26] S. Haykin, “Least-Mean-Square Algorithm,” in *Adaptive Filter Theory*, 3rd ed. NJ, USA: Prentice-Hall, 1995, ch. 9.

[27] S. Van Kuyk, W. Bastiaan Kleijn, and R. C. Hendriks, “An instrumental intelligibility metric based on information theory,” *IEEE Signal Processing Letters*, vol. 25, no. 1, pp. 115–119, Jan. 2018.

[28] S. Van Kuyk, W. Bastiaan Kleijn, and R. C. Hendriks, “An Evaluation of Intrusive Instrumental Intelligibility Metrics,” *IEEE/ACM Transactions on Audio Speech and Language Processing*, vol. 26, no. 11, pp. 2153–2166, Nov. 2018.

[29] F. S. Hallack and M. R. Petraglia, “Performance comparison of adaptive algorithms applied to acoustic echo cancelling,” *IEEE International Symposium on Industrial Electronics*, vol. II, pp. 1147–1150, Jun. 2003.

[30] R. Lee, M. S. Kang, B. H. Kim, K. H. Park, S. Q. Lee, and H. M. Park, “Sound Source Localization Based on GCC-PHAT with Diffuseness Mask in Noisy and Reverberant Environments,” *IEEE Access*, vol. 8, pp. 7373–7382, Jan. 2020.

[31] J. S. Erkelens and R. Heusdens, “Tracking of nonstationary noise based on data-driven recursive noise power estimation,” *IEEE Transactions on Audio, Speech and Language Processing*, vol. 16, no. 6, pp. 1112–1123, Aug. 2008.

[32] A. Greenberg, “Spies Can Eavesdrop by Watching a Light Bulb’s Vibrations.” wired.com. `https://www.wired.com/story/lamphone-light-bulb-vibration-spying/`. accessed Jan. 13, 2022.