CHALMERS

An Interactive Auralization Method

Auralization using ambisonics & real-time sound sources

JOSEFIN LINDEBRINK

Department of Civil and Environmental Engineering
Divison of Applied Acoustics
Vibroacoustics Group
Chalmers University of Technology
Gothenburg, Sweden 2013
Master's Thesis 2013:132

MASTER THESIS 2013:132

An Interactive Auralization Method

Auralization using ambisonics & real-time sound sources
Masters Thesis in the Masters Programme in Sound & Vibration

JOSEFIN LINDEBRINK

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Vibroacoustics Group
CHALMERS UNIVERSITY OF TECHNOLOGY

Göteborg, Sweden 2013

An Interactive Auralization Method
Auralization using ambisonics & real-time sound sources

Master's Thesis 2013:132

Department of Civil and Environmental Engineering
Division of Applied Acoustics
Vibroacoustics Group
Chalmers University of Technology
SE-412 96 Göteborg
Sweden
Telephone: + 46 (0)31 - 772 100

Cover:
Interpretation of audible environments, Martin Boyton, Stockholm,2012

Reproservice / Department of Civil and Environmental Engineering
Göteborg, Sweden 2013

An Interactive Auralization Method
Auralization using ambisonics & real-time sound sources

JOSEFIN LINDEBRINK
Department of Civil and Environmental Engineering
Divison of Applied Acoustics
Vibroacoustics Group
Chalmers University of Technology

Abstract

During recent years auralization methods have evolved towards using interactive measures. The use of interactive elements such as navigation in static sound fields has proven to be very significant in order to better integrate the listener with the simulated soundscape. In this study the possibility of engaging the user by actively contributing to the sound field is explored by enabling the subject to make sounds and communicate within the environment. Auralization allows for a subjective evaluation of the acoustical space and therefore plays an important part in a wider understanding of how we are affected by different environmental characteristics. With an auralization framework utilizing real-time sound sources direct experience of the acoustical response of the physical space is enabled and can thus be used as a tool for evaluation. Real-time convolution software implementing this method for performing auralizations has been designed. A subjective evaluation has been made using a ambisonics decoded sound field reproduced through a multi-channel loudspeaker system and a directional microphone with feedback control. Evaluation results indicate a positive response from the subjects to the added control over the simulated space. Further studies need however to be made to analyze the effects the added activity of the listener has on his or her perception of the space.

Key words : Auralization, Virtual Acoustics, Real-time convolution, Ambisonics, Acoustic feedback control, Interactive Virtual Environments, Pure Data

En interaktiv auraliseringsmetod
Auralisering med ambisonics & real-tidskällor

JOSEFIN LINDEBRINK
Institutionen för bygg- och miljöteknik
Avdelningen för Teknisk akustik
Vibroakustiksgruppen
Chalmers Tekniska Högskola

Sammanfattning

Under senare år har metoder för auralisering utvecklats till att allt mer tillämpa interaktion mellan lyssnare och ljudmiljö. Möjligheter så som navigering i statiska ljudfält har visats vara signifikanta i arbetet med att förbättra integreringen av lyssnaren i det simulerade ljudlandskapet. I detta arbete har möjligheten att låta lyssnare aktivt bidra till ljudlandskapet genom att agera ljudkälla och kommunicera i den simulerade ljudmiljön studerats. Auralisering möjliggör en subjektiv bedömning av ljudmiljöer och är därmed en viktig del i arbetet med att studera hur akustiska egenskaper hos fysikaliska miljöer påverkar oss. Med ett ramverk för auralisering som tillämpar real-tidsljudkällor tillåts en direkt upplevelse av miljöns akustiska återkoppling som i sin tur skulle kunna användas som ett verktyg i en bedömning av dess egenskaper. Initiella studier har genomförts för att fastställa eventuella fördelar och nackdelar med en sådan metod. Subjekt i genomförda studier har reagerat positivt till den adderade kontrollen av vad som hörs och när. Fortsatta studier krävs dock för att fullständigt utreda metodens betydelse när det gäller genomförandet av auraliseringar i syfte att bedömma akustiska egenskaper hos fysikaliska miljöer.

Nyckelord : Auralisering, interaktiva ljudmiljöer, real-tids faltning, real-tids ljudkällor, ambisonics, akustisk rundgångskontroll, pure data

Acknowledgements

First and foremost I would like to thank my supervisor at Chalmers University of Technology, Associate Professor Jens Forssén for all the great advice and support during the course of this work.

I would also like to thank the people at the acoustics department of Tyréns for all the input and support when working on this project, especially Philip Zalyaletdinov and Martin Höjer for discussions had during this time.

Penny Bergman, Mendel Kleiner and Erkin Asutay for all their advice and expertise, as well as the rest of the staff at the division of Applied Acoustics, Chalmers University of Technology.

Peter Lundén at SP Technical Research Institute of Sweden for advice on real-time audio processing and the Pure Data software. Pontus Larsson at Volvo Technology for advice on real-time audio processing and reproduction systems and Bengt-Inge Dahlenbäck, founder and creator of CATT-Acoustics for input on interactive auralization methods and help with the CATT-Acoustics software.

Niklas Billström, Ricardo Atienza and Björn Hellström at Konstfack for allowing access to the Sound Design Lab as well as help with the setup.

An important thank you should also be given to family and friends for endless support during the work on this thesis.

Josefin Lindebrink, Stockholm, May 2012

Notations

$L_{p,rec}$	Sound pressure level at receiver	dB
L_w	Sound power level	dB
$Q(\theta,\varphi)$	Angular direcitivity	
r	Radius	m
A'	Total absorption area	m^2S
V	Volume	m^3
ITDG	Initial time delay gap	s
t	Time	s
RT_{60}	Reverberation time	s
$H_{(w)}$	Transfer function	s
q	Gain factor	
f_s	Sampling frequency	Hz
B	Bandwidth	Hz

Contents

1 Introduction **1**
- 1.1 Background . . . 1
- 1.2 Purpose . . . 2
- 1.3 Problem description . . . 2
- 1.4 Limitations . . . 3

2 Auralization **4**
- 2.1 The concept of auralization . . . 4
- 2.2 Auralization in the acoustical design process . . . 4

3 Room Acoustics **5**
- 3.1 Room acoustical parameters . . . 5
- 3.2 Retrieving a RIR . . . 7
  - 3.2.1 Measuring a RIR . . . 7
  - 3.2.2 Modelling a RIR . . . 7

4 Methods for Performing Auralizations **8**
- 4.1 Sound source . . . 8
- 4.2 Listening room . . . 8
- 4.3 Reproduction . . . 8
- 4.4 Interactive auralization . . . 9

5 The Acoustic Feedback Problem and Measures of Control **11**
- 5.1 Acoustic feedback . . . 11
  - 5.1.1 The acoustical path – predicting the acoustic feedback problem . . . 12
  - 5.1.2 Criteria for system stability . . . 13
  - 5.1.3 Predicting maximum stable gain from reverberation time and system bandwidth . . . 13
  - 5.1.4 Acoustic echo control . . . 14
  - 5.1.5 Acoustic feedback control . . . 14

5.1.6 The multichannel case . . . 14
5.2 Modelling and measurement approach . . . 15
5.2.1 The single-channel case . . . 15
5.2.2 The multi-channel case . . . 19
5.3 Analysis of the acoustic feedback problem . . . 20
5.3.1 The single-channel case . . . 20
5.3.2 The multi-channel case . . . 20

6 FFT-based Block Convolution **21**
6.1 Concept of blocked FFT-convolution . . . 21
6.2 Time delay of convolution . . . 22
6.3 Results from time delay measurements . . . 23

7 Ambisonics **24**
7.1 The B-format . . . 24
7.1.1 Encoding and decoding ambisonics . . . 25

8 Application Architecture **26**

9 Psychoacoustic Evaluation **28**
9.1 Criteria for sound field evaluation . . . 28
9.2 Test approach . . . 28
9.3 Results of the evaluation test . . . 32
9.3.1 Part 1 - Evaluating room size using only real-time synthesis . . . 32
9.3.2 Part 2 - Comparing real-time and pre-convolved sound sources . . . 32

10 Discussion **34**
10.1 Design criteria for the auralization framework . . . 34
10.2 Reproduction level and the acoustic feedback problem . . . 35
10.3 The subjective evaluation . . . 35

11 Conclusion **37**
11.1 the Auralization framework . . . 37
11.2 Choice of equipment and acoustic feedback problem . . . 37
11.3 Short notes on a transportable auralization set-up . . . 38

12 Future Work **39**

References **41**

A Measurement Set-Up, Feedback Measurements in the Semi-Anechoic Chamber **43**
A.1 Equipment data . . . 43
A.2 Microphone directivity data . . . 44

A.3 Measurement set-up 45
A.4 Measurement results 46

B Measurement Set-Up, Feedback Measurements in the Sound Design Lab 49
B.1 Equipment data 49
B.2 Measurement results Konstfack, multichannel loudspeaker set-up 52

C Listening Test Questionnaire 54

D Auralization Set-Up 57

CHAPTER 1

Introduction

1.1 Background

In acoustical design numerous parameters help us evaluate the acoustical qualities of physical space. However, for a complete assessment perception-based data should be included. For this the sound field needs to actually be heard. Therefor studies have been made on how to recreate the acoustical ques of physical space through modelling and measurements.

Beginning with pioneering work performed by Spandöck et al [1], this early work meant measuring in physical scale models. Today, as computers have become much more powerful, calculations are primarily performed using digital 3D models. Dedicated software such as Odeon or CATT-Acoustics has been made available for calculation of the objective data of room models as well as generating audio for auralization purposes. Efficiency of testing and changing acoustical parameters has also gained through the use of digital computer models. Today the main objectives are to optimize the auralization practice through more efficient calculation of the impulse response, calculation of the impulse response in real-time as well as developing methods for adding listener influence on the simulated soundscape.

When it comes to presenting simulated soundscapes or virtual acoustic environments, VAE, interactive real-time auralization methods have in recent time been adopted more and more. Instead of having the listener simply be subjected to the environment through a listening experience, added control is given with the aim of merging the listener more with the soundscape. Recent interactive methods includes allowing the listener to move around in static sound fields or alter its acoustical characteristics in real-time, instantly hearing the effects [2, 3, 4]. By activating and engaging the listener more in the virtual environment the possibilities of a subjective assessment seems to be improving.

With auralization a more detailed evaluation of the qualities of a soundscape can be assessed, beyond the scope of objective data. With the indicated possibilities of utilizing

interaction, the question now arise as to what modes of interaction should be included in the auralization process, arriving at methods that further improves the conditions for subjective assessment. Possibly several modalities of interaction could be used that engages the listener in more ways than one, resembling the control benefitted from real environments.

Studies made by Appel and Beerends [5] on assessing the quality of one's own voice in telecommunication systems and enclosures suggests that experiencing the feedback or response in ways of reflected energy, aspects of the speech production are affected. Adjusting one's voice when speaking, such as increasing the loudness due to background noise (i.e. the Lombard effect) or altering the speech rate, implies that conscious or subconscious mechanisms are set to work. If these mechanisms in turn can be used to assess the conditions provided by the environment it offers a great tool to assess VAE. It also allows the subject to use its own voice as reference.

Methods of enabling musicians to play their instrument in VAE have been proposed [5, 6] as it has been found that musicians are well aware of the environment of which they play their instrument in. Musicians are able to alter how they play their instrument depending on the characteristics of the environment. Similarly, the pre-existing knowledge of the use of our own voice can potentially be used as a reference to assess the room acoustical qualities of VAE. Knowing the sound of our own voice and being able from a very early age to adapt the way of using it depending on the surroundings, these functions could possibly be used to evaluate the acoustical qualities of a space.

1.2 Purpose

The purpose of this work is to study how to compile an auralization framework that enables real-time sound sources within VAE. The scope of this work also includes performing initial studies of the significance of such an interaction with the hypothesis that the added interaction is beneficial to the possibilities of assessment.

1.3 Problem description

The application should utilize a multi-channel loudspeaker system and a continuous microphone feed in the listening room. The application also needs to handle real-time convolution between pre-calculated impulse responses together with a sound source present at the time of auralization. There needs to be an ability to use a flexible amount of impulse responses, i.e. enable modelling of rooms with a wide variety of characteristics. There can be no excess delays as it will deteriorate the listening experience. The framework for the application should preferably be made so that expansion such as also using pre-convolved environmental sounds is made possible. Auralization of several real-time sound sources would allow communication, the system should therefore allow for several

users to participate in the virtual acoustic environment. For several participants using a 3D – surround loudspeaker system is called for. The application is compiled with adjustable parameters to allow user flexibility. Tyréns have also requested a flexible solution that allows an effective way of performing simulations outside of treated sound labs. Only short notes on how this could be done have been included in the scope of this work. These are given at the end of this report.

1.4 Limitations

The application compiled in this work is restricted to one interactive component making it possible to only study the effects of this. However, the framework is designed in a platform that allows for further development. As only the effects of real-time sound sources are studied, fixed source and receiver positions are used, which also decreases complexity. Utilizing pre-calculated impulse responses will substantially decrease the computational load so implementation could be done using a personal working station. Pre-convolved sound sources are only used in a comparative context during listening tests. The ability to use such together with the real-time sound sources is not included at this point in the auralization framework. Performing auralizations in regular rooms, without sufficient absorption is a tedious task as the acoustics of the listening room will blend with the simulated space. In this work short notes are given as on what to avoid to prevent this from happening. The aim has been to compile an auralization framework that is suitable for presenting sound fields in the acoustical design process. Necessary approximations concerning the reproduction and simulation of VAE has been undertaken when necessary.

CHAPTER 2

Auralization

2.1 The concept of auralization

Auralization is a way of re-creating or simulating the experience of an acoustic environment and thus enable subjective evaluation of sound fields. With auralizations, possible health effects, subjective preference and environment suitability can be assessed. In practice auralization is made by modelling or measuring a room's impulse response, RIR. The impulse response includes information on how sound is transmitted from a source to a receiver position in a physical space. The transmission path can later be convolved with audio and thus the transmission can be simulated. The RIR contains information on the transmission of a sound that is emitted in space. The sound will be absorbed, scattered or reflected interacting with the different physical objects or restraints of an environment.

2.2 Auralization in the acoustical design process

Since auralizations are a way of making the transmission path audible of sound emitted in an environment by a sound source to a receiver, it is a great tool to use in the acoustical design process. Auralization enables direct comparison with and without measures of control. The theoretical nature of traditional acoustics has made the possibilities and importance of acoustic design somewhat inaccessible to people outside of the acoustics-community. Using auralizations, immediate access can be given to experiencing possibilities of an adequate acoustical design. As Furlong et al. have discussed [4] simple solutions need to be applied, enabling an insight to the acoustical design process, allowing this to be considered at an early designing stage.

CHAPTER 3

Room Acoustics

3.1 Room acoustical parameters

Sound pressure level, $L_{p,rec}$ at the receiving point will consist of direct sound from the sound source as well as reflected sound from the enclosure. The resulting sound pressure level, Lp,rec , is dependent on the sound power level of the sound source, directivity of the sound source as well as total absorption area of the enclosure.

$$L_{p,rec} = L_w + 10\log[(Q(\theta,\varphi))/(4\pi r^2) + 4/(A')] \qquad (3.1)$$

A room impulse response, RIR, gives time-based information on the transmission of sound from a source to a receiver position. The initial time delay gap, ITDG, seen in the RIR, is a measure of the time between the first incoming sound at t_0, and the first reflection, occurring at time t_1, depicted in figure 3.1. If this gap is long there is a risk of echoes occurring.

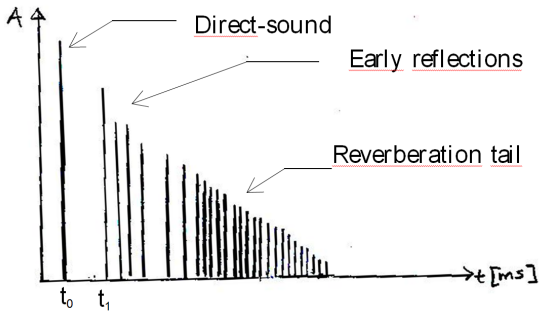

Figure 3.1: Example of an impulse response showing the direct sound, early reflections and reverberation tail

Echoes occur when reflections of a certain sound pressure level arrive at the listener with a sufficient time delay to the direct sound. For speech applications echoes can be

detrimental as the talker is disrupted by retardant information. The nature of speech makes us more sensitive to echoes occurring. Our sensitivity to echoes is dependent on speech rate, angle of reflection incidence as well as reflection ratio to direct sound concerning sound pressure level and time delay. Tests conducted by Haas, figure 3.2, show that with an equal sound pressure level, the annoyance or the detection of an echo was reported by 50 % of the test subjects at a delay time of about 60 ms when subjected to two sounds with the same incidence angle. From these tests one can also see indications that around 50 ms, the delay of a reflection with a -10 dB lower sound pressure level to the direct sound would start being detectable, indicating a threshold for echo detection. Many spatially dependent parameters utilize the time threshold of 50 ms to distinguish early sound information from late. Haas tests indicate that a 50 ms time difference starts to have importance when the sound pressure level difference is about 10 dB.

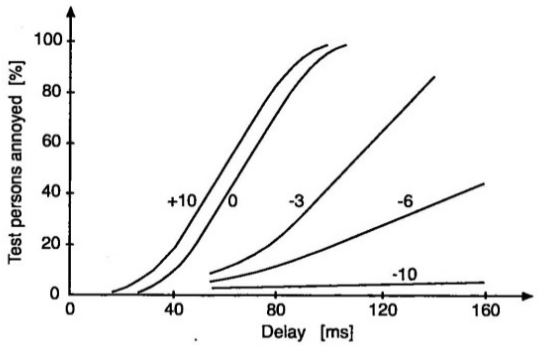

Figure 3.2: Results from Haas measurements of echo annoyance between two signals. Results show dependence on sound pressure level difference as well as with different time delays. Both with a 0 degree angle of incidence [7]

Reverberation time, or reverberance, often abbreviated RT_{60} measures the time from steady state sound pressure level in diffuse field to the time it decreases by 60 dB as the sound source is switched off. The reverberation time can be estimated using Sabine's formula:

$$RT_{60} = 0.161 \frac{V}{A'}, \qquad (3.2)$$

Where V is the volume of the enclosure and A' is the total absorption area. Reverberation time gives a sense of room size and is one of the most fundamental room acoustical parameters for subjective perception [8].

Coloration of the sound field is unwanted changes in the frequency spectrum either caused by non-linearity's in the recording or reproduction system or by room reflections.

Extensive coloration causes deterioration of sound quality.

3.2 Retrieving a RIR

RIR is dependent on source and receiver positions, source characteristics as well as geometry of the enclosure and materialistic properties of spatial boundaries. The RIR needs to be adapted to the chosen reproduction method. That is, for binaural reproduction two channels are produced corresponding to the right and left ear signals when using headphones utilizing Head-Related Transfer Functions, HRTF. For multichannel loudspeaker set-ups, the 3D sound field reproduced by the loudspeakers needs to be encoded. Most commonly the ambisonics technique is utilized resulting in a 4-channel RIR covering the coordinates of the physical space.

3.2.1 Measuring a RIR

Different methods for measuring RIRs are available such as the Maximum Length Sequence, MLS, the Inverse Repeated Sequence, IRS, Time-Stretched Pulses or sine sweep. For a multi-channel loudspeaker reproduction the RIR should be measured using an Omni-sound source combined with a microphone containing multiple capsules for incidence angle. For binaural reproduction a binaural dummy-head can be used. When measuring RIRs it is important to keep the electronic SNR low. All combinations of the source and receiver positions need to be recorded separately.

3.2.2 Modelling a RIR

Modelling and calculating a room impulse response, the geometry and material data must be defined. Calculations usually include image-source modelling of the early order reflections combined with calculation methods such as ray-tracing for the later order reflections, including the reverberation tail. Room boundaries need to be defined as well as surface and material properties. Absorption and scattering coefficients are set for each surface. The sound source is defined with sound power level, directivity properties, aim and position. The receiver is defined with position and head direction. The receiver characteristics also needs to be specified from the reproduction method chosen to derive a suitable RIR.

CHAPTER 4

Methods for Performing Auralizations

4.1 Sound source

For auralization purposes, audio needs to be convolved with the impulse response. The sound source needs to be free of any room influence and should preferably be recorded in an anechoic room or in a highly damped environment.

4.2 Listening room

For loudspeaker reproduction the listening room needs to be free of any distinct room influence. The total absorption should be high in a broad range of frequencies. Early reflections from the listening room should be avoided and the reverberation time should be kept sufficiently shorter than that of the simulated environment's. Essentially the listening room sets a limit for which rooms can be simulated.

4.3 Reproduction

Most commonly a binaural setup or multiple loudspeakers are used for reproduction of acoustic simulations. With headphones the listening room has less impact on the simulation but the binaural signals need HRTF-filtering. When using a loudspeaker system there needs to be sufficient loudspeakers so that the 3D-acoustical cues are reproduced at an appropriate detailed level. The 3D-image needs to be kept intact. Most commonly ambisonics or in some cases wave-field synthesis is used to decode the sound field when utilizing loudspeakers.

4.4 Interactive auralization

For a real-time interactive auralization application proposed in this thesis the source and receiver would be one and the same. Therefor the source and receiver distance should theoretically be the distance of a person's mouth to his or her ears. The reproduced sound field should not contain the direct sound of the talker as this would be present at the time of auralization. As the source signal will be generated at the time of auralization, the direct source signal of the RIR, shown in figure 4.1, should be edited out prior to auralization.

A microphone needs to be added within the listening room to retrieve the source signal. This signal needs to be continuously feed to the convolution algorithm as the convolution needs to occur in real-time. The source signal is convolved with the edited RIR containing only reflected energy of the simulated environment. The calculated sound pressure level difference between direct energy and reflected sound energy needs to be kept intact at the point of the receiver in the listening room. Therefore the system's reproduction level needs to be calibrated so that the corresponding sound pressure level difference is achieved. As the sound power level of the source is changing during the course of auralization this ratio needs to be intact and the reproduction level a consequence of the dynamic change of the sound source power level.
The ratio between direct sound and reflected sound would be varying between different RIR's, thus the system needs to be calibrated for each RIR. If normalizing the RIR's too each other, the same reproduction level could theoretically be used. Modelling the source and receiver at such close distance as the distance between mouth and ear, approximately 0.10-0.15 m, can create problems. An average standing height of 1.7 m means the travelling distance is about 34 times longer for one of the earlier reflection to reach the receiver than the direct sound. A sufficient dynamical range is needed to avoid electronic or digital noise. To model the heads effect on the direct sound transmission is also a tedious task and approximations are called for.

Excess time delays should be suppressed so that no audible echo, unnatural to the

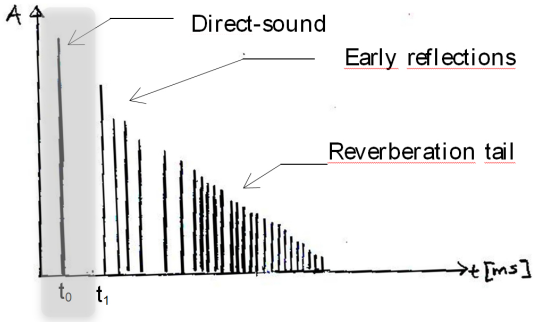

Figure 4.1: Illustration, editing of RIR.

simulated environment occurs. The time of processing needs to be kept as short as possible.The time it takes from the sound being produced by the talker, runned through the application and reproduced at the listener's ear should not exceed the ITDG of the RIR beyond the limit of audible delay, that is 20-30 ms. [5]

CHAPTER 5

The Acoustic Feedback Problem and Measures of Control

5.1 Acoustic feedback

Acoustic feedback is a problem one usually has to address when dealing with electro-acoustic systems. Having chosen to use a multichannel loudspeaker system coupled with a live microphone placed within the same enclosure, this is likely to occur as the loudspeaker signal can re-enter the system through the microphone. Transmission through reflection from enclosure boundaries makes this even more likely to occur. The acoustic feedback phenomenon can manifest through single narrow-band oscillations giving rise to self-oscillation or a so-called howling effect. This is referred to acoustical feedback. When a broader band of the signal re-enters the microphone one instead talk about acoustic echo. Depending on the nature of the problem, different measures of control are necessary to adapt. With a single-channel case there is a forward- and feedback path (figure 5.1-5.2). The forward path, *G(w)*, will be defined as the electro-acoustical path from the microphone to the loudspeaker, transmitted through the enclosure to the listener. Here the microphone signal undergoes processing determined by the user, is amplified with a frequency independent gain factor q, and sent to the loudspeaker. The feedback path, *F(w)*, connects the loudspeaker output to the microphone, creating a closed-loop system [9].

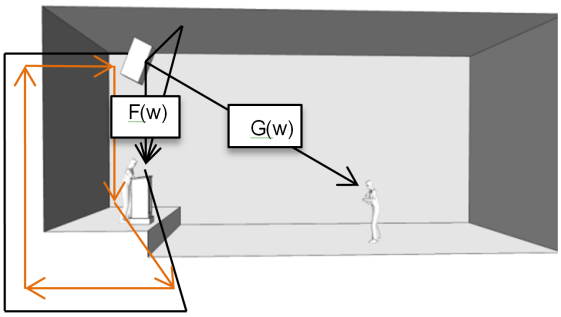

Figure 5.1: Schematic of the acoustical feedback problem showing the forward and feedback path of a single-channel case.

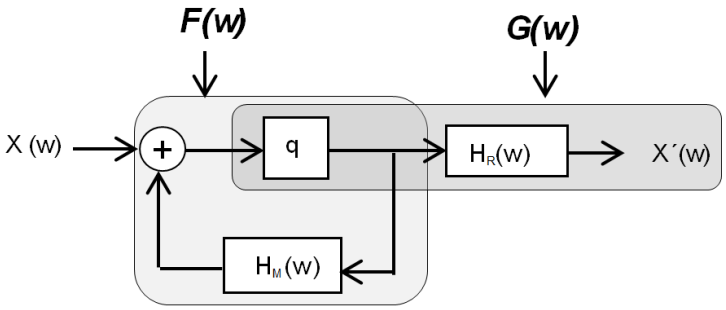

Figure 5.2: The forward and feedback path

This being a highly simplified case, not dealing with the transfer functions of the electronic system components and also not the specific nature of room boundary reflections and their impact on the transmission of the closed loop system. It is also a highly simplified form of describing the transfer functions of the enclosure, that is $H_{R(w)}$ and $H_{M(w)}$.

5.1.1 The acoustical path – predicting the acoustic feedback problem

The acoustic feedback problem is hard to predict as it is a consequence of many factors. The transfer paths within the enclosure are subjected to change from people's movement as well as air conditions. This makes it hard to foresee and difficult to predict. Expanding the problem to a multi-transmissional case it gets even more complex to predict. Several ways of explaining and dealing with acoustic feedback and acoustic echo issues have been proposed, one of them being the Hänsler and Schmidt's LEM-model [10] including the effects of the enclosure.
The Loudspeaker- enclosure – microphone system, (LEM), take the positions of the different elements within in the enclosure in to account. If the main loudspeaker transmission lobes are directed away from the microphone, the LEM system mainly connects

through reflections from room boundaries, or off subjects and objects in the room, so called indirect coupling. Direct coupling occurs when the loudspeaker signal is directly transmitted to the microphone. Studying the single-channel case in figure 5.1, the input spectrum will be that produced by the talker at the point of the microphone, X(w). The signal is sent through the electro-acoustic forward path, G(w), with processing and amplification by a factor q. Two transfer paths are of interest, from the loudspeaker system to the receiver position and to the microphone,H_M. The resulting signal spectrum at the receiver, X'(w), is determined using:

$$X'(w) = qH_{R(w)}[X_{(w)} + H_{M(w)}\frac{(X'_{(w)})}{(H_{R(w)})}], \tag{5.1}$$

Where q is the gain factor, $H_{R(w)}$ is the transfer function between loudspeaker and receiver and $H_{M(w)}$ is the transfer function between the loudspeaker and microphone. Equation 5.2 gives us the resulting transfer function, $H_{tot(w)}$of the entire system:

$$H_{tot(w)} = \frac{Out}{In} = \frac{X'_{(w)}}{X_{(w)}} = \frac{qH_{R(w)}}{(1-qH_{M(w)})} = qH_{R(w)}\sum_{(n=0)}^{\infty}[qH_{M(w)}]^n$$

The factor $qH_{M(w)}$ is characteristic for the amount of feedback and is called the open-loop gain factor. This factor dominates the difference between the input spectrum $X_{(w)}$ and the resulting spectrum at the listener $X'_{(w)}$. For indirect-coupled acoustic feedback, Schroeder explained that the usual behaviour of the feedback is peaks at the favoured frequencies of about 10 dB higher than the average sound pressure level, spaced about 10 Hz apart. [12]

5.1.2 Criteria for system stability

To ensure a stable system one can use Nyqvist's criteria which say that if for an angular frequency of w, the following statements are true:

$$\begin{cases} \mid (qH_{M(w)} \mid \geq 1 \\ \angle qH_{M(w)} = n2\pi \end{cases}, n = 1,2,3.....(5.3)$$

the system is likely to become unstable and audible ringing will probably occur.

5.1.3 Predicting maximum stable gain from reverberation time and system bandwidth

In cases where $G_{(w)}$ has a relatively smooth magnitude response, and the Schroeder condition is fulfilled, i.e. the coupling between loudspeaker and microphone is mainly due to reflections of the enclosure, Schroeder developed a criterion for *maximum stable gain, MSG* also called *gain before instability (GBI).* With a known system bandwidth B and a reverberation time of RT_{60} the *MSG* can be calculated using equation 5.4. [11]

$$MSG(t) = -10\log[\log(\frac{BRT_{60}}{22})] - 3.8,\ \text{dB} \qquad (5.4)$$

The *MSG* states the maximum, i.e. the highest possible amplification before feedback is a definite problem. However, a margin of safety is important to ensure a stable system throughout use, especially when possibly having moving subjects in the listening room. For speech applications, it is recommended to use a margin of about 5 dB to the MSG. [12]

$$20\log(\frac{q}{q_0}) \leq -5\ \text{dB} \qquad (5.5)$$

Where:
q_0: Critical value of the gain factor for which the system goes unstable
q: Gain factor used

5.1.4 Acoustic echo control

Acoustic echo control is needed whenever a broad-band portion of the loudspeaker spectrum is transferred back into the microphone of any active system, not fulfilling the Nyqvist's criteria. This is usually done using adaptive filtering like the least mean square, LMS - method. The basic idea is to record the source signal with a control microphone and use the inverted frequency spectrum of the recorded signal as a filter to cancel out the unwanted sound. Different echo-cancellation systems have been compiled for telecommunication conference systems. These are however based upon the sending and receiving room being two separate.

5.1.5 Acoustic feedback control

If only acoustic feedback control is necessary, there are frequency-cancellation systems available. Many of them are based upon narrow-band filtering like the notch-filter based howling suppression, NHS. These can either be automatic, then the system itself senses frequency components starting to oscillate and cancels these, or they can be manually controlled where the filters are set either beforehand or during run-time. For the auralization application, an automatic system is favourable. The system needs to sense oscillating frequencies in a very short time so that audible ringing does not occur. The system also needs to have a sufficient amount of filters, to handle multiple feedback frequencies, especially when going from a single-channel to the multichannel case.

5.1.6 The multichannel case

Expanding the reproduction system from one single channel to multiple channels, the problem complexity grows vast dealing with multiple transfer paths, both direct and indirect as well as multiple transfer functions. However, the NHS-method can still be

used as long as the problem is restricted to only narrow-band frequency oscillation. Since loudspeakers are placed at different positions to the microphone, there are possibilities of more feedback frequencies and additional filters are therefore necessary.

5.2 Modelling and measurement approach

To ensure a stable system, direct or indirect coupling needs to be avoided. Thus the sound pressure level of the sound source, i.e. the subject, should be high in comparison to the loudspeaker signals at the point of the microphone. This is in conflict with the criteria of the auralization method, which states that the sound pressure level ratio between direct sound energy from the subject and reflected sound energy reproduced by the loudspeakers should remain intact at the point of the sound source (as well as receiver's position).. With a multichannel loudspeaker set up, the loudspeakers should not be directive as the aim is to reproduce a full 3D sound field However, for a transportable auralization set,a directive stereo system could be necessary to use as the reproduction rooms are likely to not be sufficiently treated. In both cases, narrow pick-up equipment can be used to prevent feedback from occurring. To ensure a sufficient sound source signal, the microphone should be placed as close to the talker as possible, but still keeping the microphone's presence negligible to the talker.

A model of the acoustical feedback problem has been made and validation measurements performed to study what suppression of acoustic feedback can be achieved for the auralization application. To start off, a single-channel case is studied, later expanded to a multi-channel case. A model of the acoustical feedback problem, using the single channel case was made using the CATT-Acoustics software and different directional characteristics of the recording and reproduction system. The results of the models were then validated through measurements in a semi-anechoic chamber. Finally a multi-channel case was measured in a treated sound lab, the Sound Design Lab at the at the University College of Arts, Crafts and Design.

5.2.1 The single-channel case

Measuring in a semi-anechoic chamber, room boundary reflections except from a hard floor surface can be neglected, studying only the effects of directivity and positioning of the equipment as well as direct coupling together with a controlled indirect coupling. The subject is for measurements replaced with an omni-directional loudspeaker referred to the talker to avoid any confusion. Positions of the loudspeaker and microphone were chosen so that their main lobes where directed towards the talker and away from each other, figure 5.3 shows early sketches of this. In configuration 1, the loudspeaker is placed facing the talker with a 0 degree vertical and horizontal angle and the microphone in a 45 degree vertical and 0 degree horizontal angle. Two different distances between

loudspeaker and talker were used, 2 and 3 m. The microphone was placed at distances 0.2, 0.4, 0.8 and 1.6 m from the talker. In configuration 2, the loudspeaker was instead hanging from the ceiling directly above the talker. The microphone was placed in the same way as in configuration 1. The derived transfer functions where then compared directivity and position- wise.

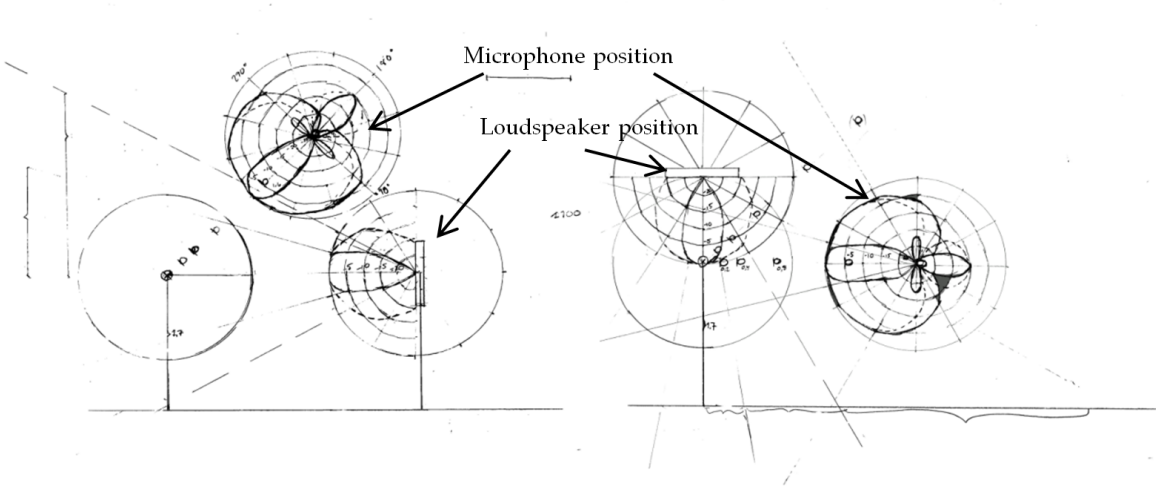

Figure 5.3: Early sketch of the positioning and directivity of equipment used in the model and for the measurements.

Model of the single-channel case

A model of the semi-anechoic chamber was made using Google Sketchup and calculations were made using the CATT-Acoustics software. The model is shown in figure5.4 and 5.5. For reproduction both a line source and an omni-directional source were tested. The talker and loudspeaker sources were given similar directivity data as the loudspeaker used for measurements. Directivity data for the loudspeakers used for measurements were not available so data on similar loudspeakers was in the end used for the model. The talker source was defined using a human voice directivity as the loudspeaker used for measurements was not determined at the time of modelling. Source directivity is depicted in figure 5.6. The sound pressure level of all sound sources were calibrated to each other at 1 m distance before calculations for comparison reasons.

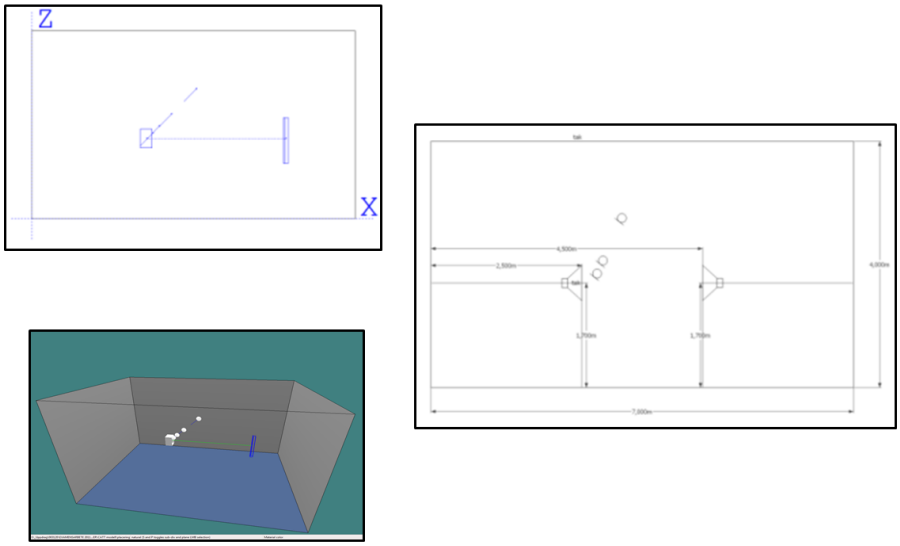

Figure 5.4: Configuration 1, loudspeaker placed in front of talker.

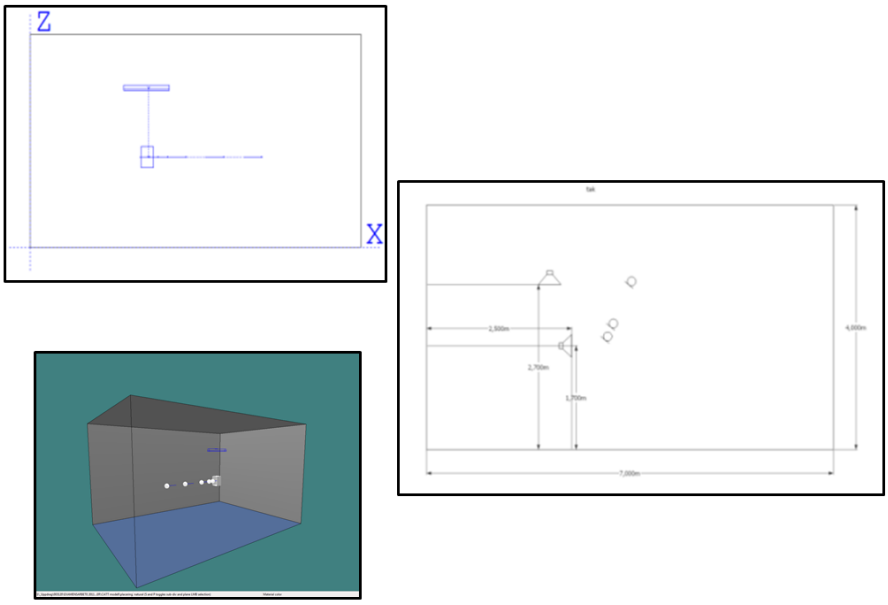

Figure 5.5: Configuration 2, Loudspeaker hanging above the talker.

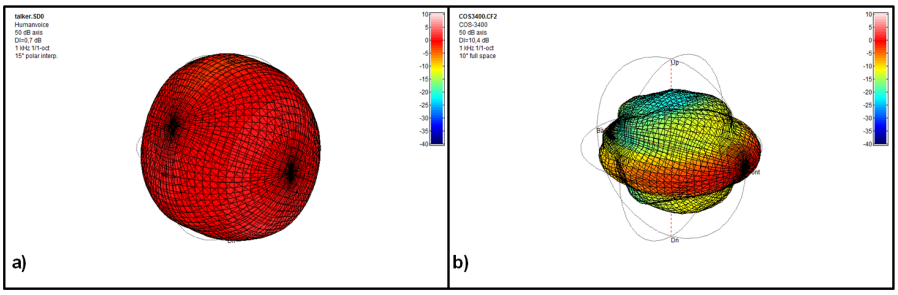

Figure 5.6: Directivity plots at 1000Hz for the a) talker source and b) line source used in the model

Different microphone directivities were used, OMNI, Cardioid, Super-Cardioid and Shotgun. The shotgun directivity had to be manually specified using equation 5.6.

$$F = Z + X\cos\theta (5.6)$$

Where X defines the sensitivity on the main axis that is at a 0 degree angle, Z defines the sensitivity at an angle, θ. For OMNI directivity, X is set to 0 and Z to 1, for Cardioid directivity, X =1 and Z =0. In these tests, X=0.85 and Z=0.15 were used to define the shotgun directivity. Calculated transfer functions between the talker and microphone where compared to the derived transfer functions between loudspeaker and microphone.

Measurements of the single-channel case

The semi-anechoic chamber at Tyréns was used for validation measurements, with the walls and ceiling of this chamber covered with wedged mineral wool, and the floor surface made of heavy untreated concretes figure 5.7. The same configurations used in the model were also applied, having the talker source replaced with an OMNI-directional loudspeaker for comparison purposes.

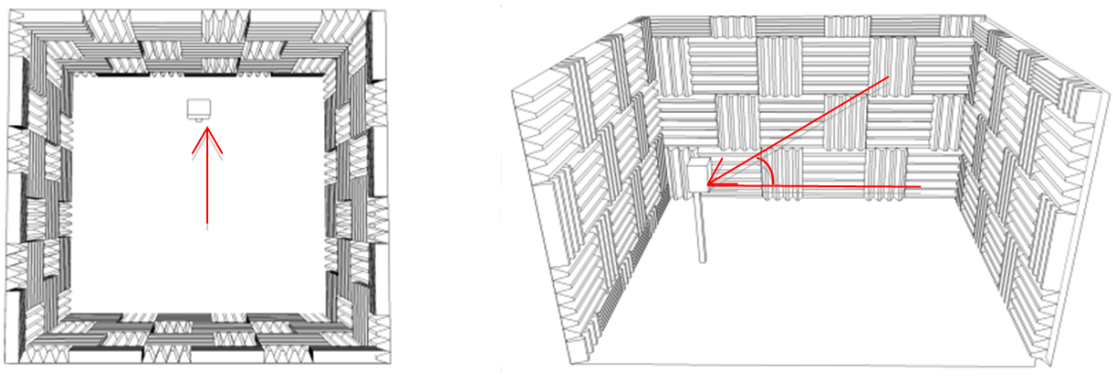

Figure 5.7: Schematic of the semi-anechoic chamber. Red arrows indicate the angle and direction of the microphone.

Again the transfer function between talker and microphone, and loudspeaker and microphone was derived. These were measured one at a time, using a sine sweep as a source signal and in such an order so that measuring positions stayed the same.

5.2.2 The multi-channel case

Measurements were carried out in the Sound Design Lab where the application is being implemented. The room is treated with absorbers and diffusors having a very short reverberation time, around 0.20-0.25 s. A large volume above the ceiling of the room controls the low frequency modes. In the room a 5.1 channel system is installed utilizing a total of 9 mid- and high-range loudspeakers, three in the front and sets of three surround loudspeakers on each side. Four subwoofers are also installed. The room is furnished with a control desk and an audience seating arranged in the back of the room.
There was two possibilities of installing a microphone in the room one having it suspended from the ceiling at an approximate 2 m distance from the audience seating at a 45 degree angle (configuration 1), or having it placed on the control desk at a same angle to the audience area (configuration 2). Configuration 1 would mean placing the microphone outside of the main horizontal axis of the loudspeaker system but further away from the sound source. The opposite conditions were the case when testing configuration 2. Since the loudspeaker calibration had to be disconnected, calibration was made ensuring the same reproduction level at the receiver point. The talker source gain was set sufficiently high and remained the same for each measurement. Again the transfer functions were measured one at a time.

5.3 Analysis of the acoustic feedback problem

5.3.1 The single-channel case

The shotgun directivity showed the worst results, i.e. less differences in sound pressure level between the loudspeaker and the talker, at the larger tested distances than the other microphone directivities. This could possibly be due to the back-lobe of highly directive microphones. However at closer distances to the talker the shotgun directivity showed an average 10 dB difference over the lower tested frequency spectrum (250-500Hz) and an average 15 dB difference in the upper tested spectrum (1000-2000Hz) between the transfer functions, giving the best results between the different microphone directivities. Possibly due to the floor reflection, having the loudspeaker hung from the ceiling gave less difference in sound pressure level between the different transfer functions. Test showed that for the single-channel case, a shotgun microphone at close distance to the talker should be used in combination with a loudspeaker placed in front of the talker. The different loudspeaker directivities showed smaller differences.

5.3.2 The multi-channel case

Measurements of the multi-channel case showed best results when placing the microphone close to the sound source within the main axis range of the loudspeakers. Audible ringing was proven difficult to avoid for this case as multiple narrow band frequency oscillations occurred frequently. Tests indicate that the microphone should be placed even closer to the sound source.

CHAPTER 6

FFT-based Block Convolution

6.1 Concept of blocked FFT-convolution

Convolution is a method of 'adding' two signals where an input signal is processed by a filter kernel. With auralization the room impulse response, RIR, is used to filter the source signal and so the sound source is 'placed' at the point defined in the model of the physical space. There are different ways of performing this convolution and many algorithms to choose from. For real-time audio processing the computational time and computational load are of highest importance. Most algorithms today have sufficiently low time delays and with hardware available, processing is more powerful. However, with long filter kernels as in the case of RIR's, the execution scheme of the convolution needs to be efficient. As can be seen from eq. 6.1, RIR-filters are very long even those derived from small and damped rooms. Thus the RIR should be divided into segments before processing. The segments should be processed in parallel to keep the execution time short.All signal processing of the application together with delays caused by acoustic transmission paths all need to sum up to an execution time below that of audible delay. As the convolution is one of the most time-consuming efforts have been made to optimize the execution of this.

$$Blocksize = t \times f_s = 0.5 \times 44100 = 22050 \text{ samples}, \tag{6.1}$$

Blocked Fast Fourier Transform - convolution, (FFT-convolution), uses the overlap-add method to divide the convolution filter (i.e. the RIR) into segments or blocks. To be able to use the FFT-convolution the blocksize needs to correspond to 2^N. FFT-convolution is much more time effective than Discrete Fourier Transform for signals longer than about 40-60 samples. A schematic of the convolution process using the overlap-add method is shown in figure 6.1.

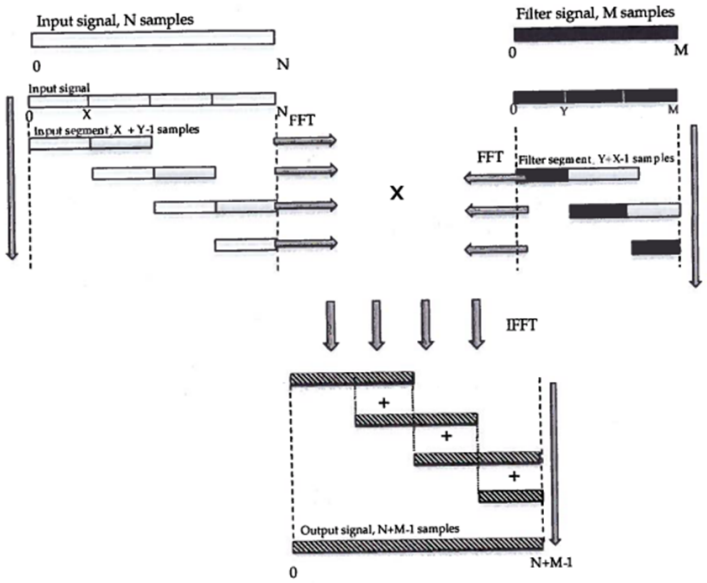

Figure 6.1: Convolution processes of an input signal and a filter using FFT-convolution and the overlap-add method.

After dividing the filter and input signal into sections, the segments are transformed with FFT to real and imaginary part frequency spectrums in the frequency domain. In the frequency domain multiplication occurs which is the equivalent to convolution in the time domain. The resulting output segments are transformed back into the time domain using inverse-FFT. Here the output segments are then re-combined into the resulting output signal.

6.2 Time delay of convolution

Since dealing with long room impulse response filters, the main time consumer in the application will be the convolution between the microphone feed and the room impulse response. To ensure that the convolution algorithm chosen is suitable, a measurement of the time delay caused by this signal processing was measured. A convolution patch in the Pd-extended library called *partconv* $\sim$, created by Ben Saylor, proved to be sufficent to use for this purpose. The *partconv* $\sim$ uses the FFTW library which contains several algorithms for performing blocked fast Fourier transforms. Depending on available hardware a suitable algorithm for computing the FFT-convolution is chosen. [13]With a user defined block-size the convolution process could be optimized to the length of the impulse response used. The convolution was tested in-software using a normalized sine sweep of 0.4 s/octave as sound source. By sending the sine sweep from a wav-file player directly to a wav-recorder as well as through the *partconv* $\sim$-object to another wav-recorder, the time difference between these processes would give the time delay of the convolution.

6.3 Results from time delay measurements

The time delay proved to be correlated to the size of one block. Thus a blocksize of 512 samples and a sampling frequency of 44.1 kHz would result in a time delay of:

$$\frac{512}{44100} = 11{,}6 \text{ ms}, \tag{6.2}$$

which is below the audible limit and also gives some extra headroom for additional processing delays. One limitation of using this convolution algorithm is that the input signals can only be divided into maximum 256 partitions i.e. blocks, of 2^N size, which limits the total length of the RIR to about 3 s when using a block size of 512 samples. This is deemed acceptable for the application at this stage. When using shorter RIRs, the process could be further optimized by lowering the segment size to 256 samples. This would result in a much more time efficient convolution process giving only a delay time of 5.8 ms.

CHAPTER 7

Ambisonics

Ambisonics is a method of encoding and decoding a 3D sound field. It allows for great variability of loudspeaker set-ups. The sound field is recorded in a single point and the sound field encoded in different channels depending on angle of incidence. Microphones dedicated to the ambisonics technique are built out of several membranes directed in a pattern shown in fig 7.1. The microphone utilizes small membranes for point reception. Ambisonics allows for any number of loudspeakers to be used although a sufficient amount is necessary to reproduce a complete sound field and to provide necessary localization ques. The loudspeaker formation is not fixed although the loudspeaker set-up should make sure there are no holes in the auditory image. Different orders of ambisonics can be used depending on the number of input channels. The 1th order ambisonics is the basic ambisonics format containing 4 channels, the so-called B-format.

7.1 The B-format

The B-format is a 4 channel recording method, where sound is recorded in a single point. The 4 channels, W,X,Y, Z, contains information depending of angle of incidence, W being the OMNI input, an average of all incidence angles, X contains information on sound arriving from left and right, Y the front and back and finally Z which handles height information. The four channel input, handling different angles of incidence is a consequence of the microphone membrane placement. From these four channels one can reproduce an entire sound field. Often sound reproduction in 2D, i.e. only the horizontal field is used, thus the Z-channel can be excluded. This method many believe still gives an accurate depiction of the sound field since often humans have a hard time discriminating sounds in the median plane.

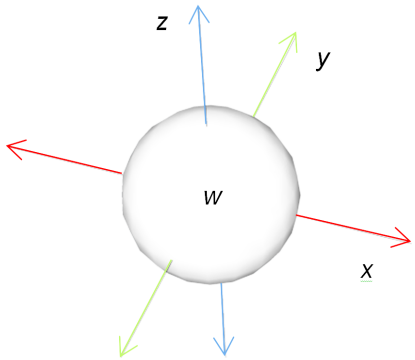

Figure 7.1: Ambisonics B-format pick-up pattern.

7.1.1 Encoding and decoding ambisonics

Most ambisonics recording microphones first registers the incoming sound field in what is called the A-format. The recording then needs to be encoded to the B-format. This is either done by dedicated hardware whilst recording or in software applications. From the B-format, the recording then needs to be decoded to the reproduction system. One advantage of Ambisonics is that any number of loudspeakers can be used, however, a sufficient amount is necessary to be able to reproduce a uniform sound field, and still, the more loudspeakers used, the more details of the sound field can be reproduced. The B-format is decoded to the loudspeakers depending on their angular position to a reference point. The loudspeaker will receive a certain amount of each B-format channel determined by a scaling matrix compiled during decoding.

CHAPTER 8

Application Architecture

The auralization application was implemented using the graphical programming software Pure Data, Pd. The application in its rough stage contains a loading function of the 4-channel RIR imported as combined wav-files. The RIR together with a continuous signal from the microphone in the listening room are sent to the convolution function, *partconv* $\sim$, of the Pd extended library. An ambisonics decoder is utilized, also obtained from the Pd extended library compiled by Thomas Musil at the Institut für Elektronische Musik und Akustik in Graz, Austria. This is used for the distribution of the convolved signal to the loudspeakers, able to handle higher orders of ambisonics, producing both two- and three-dimensional sound fields, including the z-channel of the B-format. Number of loudspeakers as well as relative angle is easily set within the application. Time delay- as well as separate gain-units is applied to each loudspeaker signal, giving opportunity to calibrate the loudspeaker system in-software if necessary.

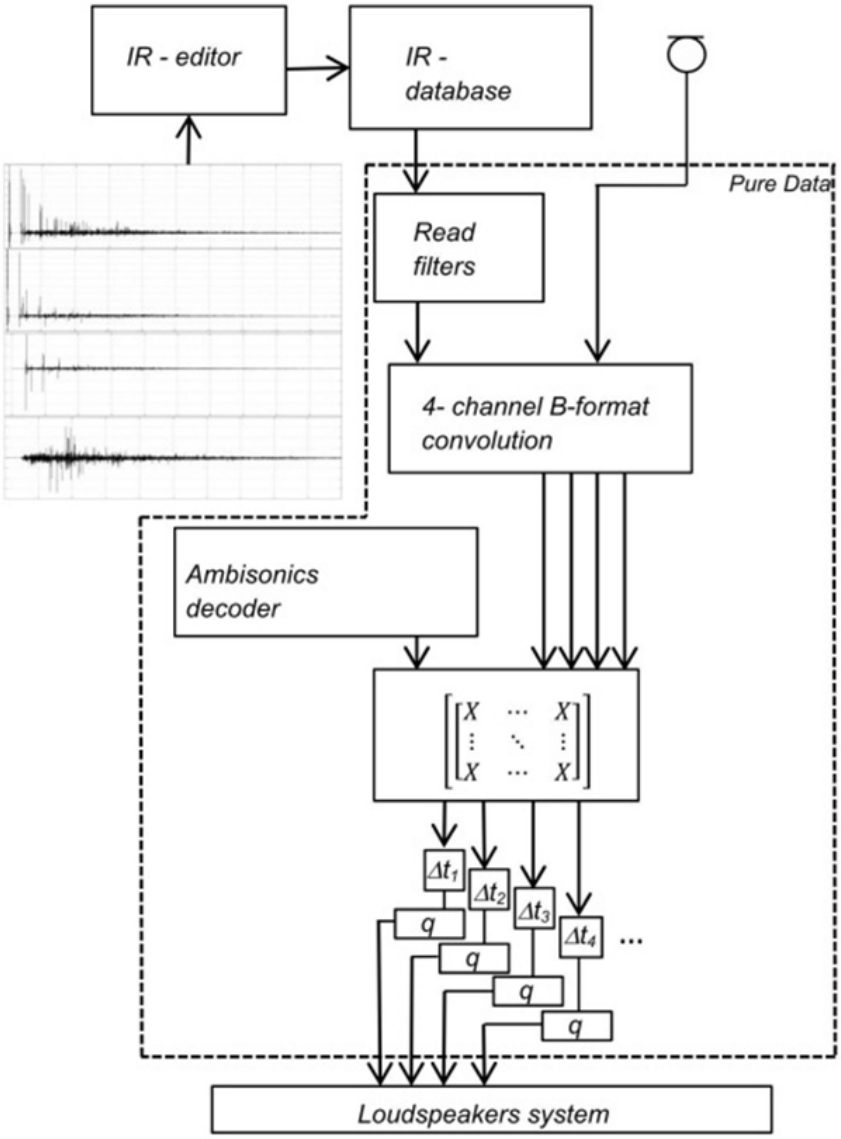

Figure 8.1: Application architecture scheme

CHAPTER 9

Psychoacoustic Evaluation

9.1 Criteria for sound field evaluation

The intention of the method proposed in this thesis is to study the effects of the added interaction by having the subject contribute as a sound source within the simulated space. Tests have been conducted seeking differences in the subject's response when exposed to environments utilizing different methods for performing auralization. As a hypothesis, the added control of the real-time sound sources application, enabling the utilization of factors such as altered speech rate and the Lombard effect, benefits the ability to perceptually judge the physical spaces acoustical qualities through the direct experience of the acoustical response. However, the added activity of the subject could also be distracting to the experience of the acoustic simulation. Thus at this initial stage of testing the preservation of fundamental acoustical qualities as well as potential benefits/restrictions of experiencing this through a direct response compared to a simple listening experience seems relevant to start with.

The tests have been conducted in a comparative context between having the subject only listen to sound events occurring during simulation to the one where real-time sound sources are utilized.The aim of the tests have not been to judge if the subject can perceive a large hard surfaced room as such but rather to see if the same subject would judge the same simulated room with the same parameters or perceive it as different using the different methods of auralization. Perceived size of rooms based on the respective RT_{60} as well as tonal character was used as a basis for these tests.

9.2 Test approach

The tests were performed using a smaller group of 9 participants in the Sound Design Lab at Konstfack, (University College of Arts, Crafts and Design). For the tests the installed multi-channel loudspeaker system was used and a directional microphone was placed on a 1 m distance from a designated seat where the subjects were asked to sit. During tests

the room was kept dark to avoid any visual influence on the results. For the study, RIRs of realistic environments were chosen with varying room geometry but trying to keep the total absorption area not to any extreme, as reverberation time are dependent on both factors. Again, usually listening tests using auralization are not conducted with such a comparative purpose and the aim is not to arrive at an absolute value for these. Therefor rooms with widely varying size and absorption were used. When evaluating the results, the rooms where categorized into large, medium sized and small rooms, as a more detailed description would for these tests be insignificant. Modelled rooms used for testing where provided by Tyréns AB. Also a measured room was used, the Great Hall of People's Palace in London, recorded by students at Centre for Digital Music, Queen Mary, University of London [14]. The latter proved to have low-frequency artifacts in its RIR, and the RT_{60} of this had to be estimated. The respective reverberation time of the environments are presented in table 9.1 and 9.2. The models used as well as a picture of the great hall is shown in figure 9.1 and a representative RIR for each environment is shown in figure 9.2.

Table 9.1: Acoustic parameters for rooms used in listening tests. The RIR of the Great hall contained low-frequency artifacts and had to be estimated.

Venue	Reverberation time, [s]
Opera Hall	1.4
Lecture Room (small auditorium)	0.4
Great Hall (measured)	2.1*
Open-plan Office	0.4
Canteen/Cafeteria	0.4

Table 9.2: Tested rooms divided into groups depending on the length of the reverberation time.

Room Size	Reverberation time, [s]
Large SizedRooms	≥ 1.4
Medium Sized Rooms	$0.5 - 1.4$
Small Sized Rooms	≤ 0.5

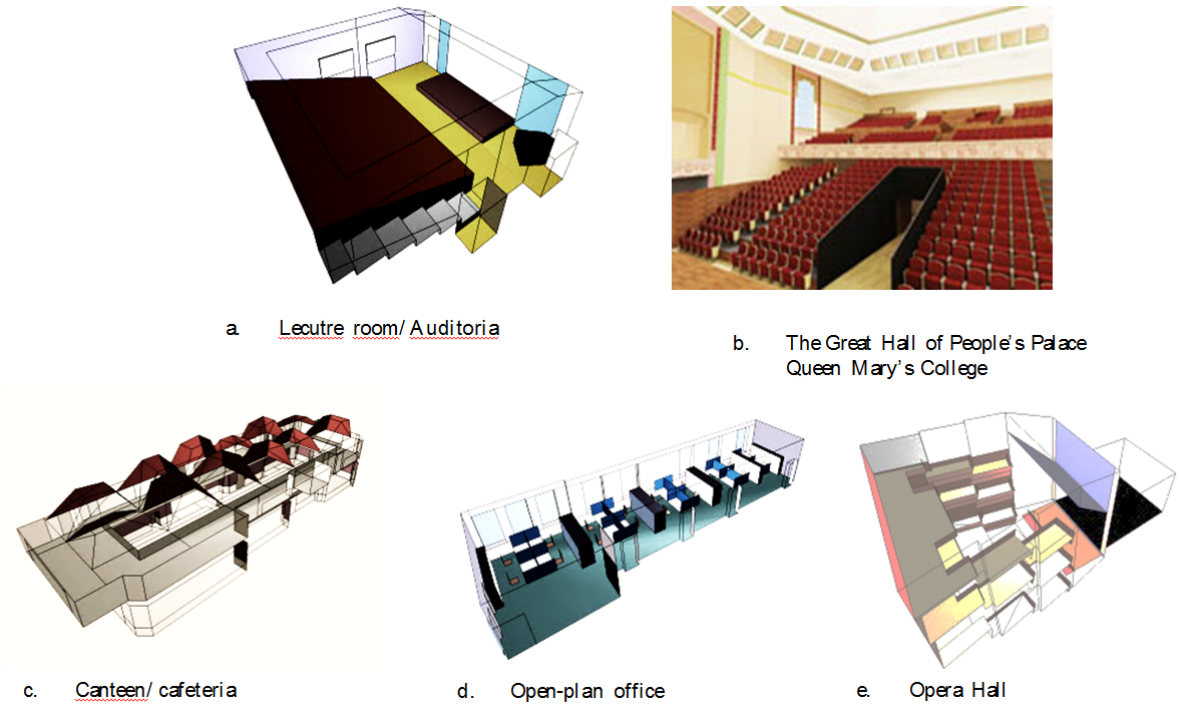

Figure 9.1: a)-e) - Models of the rooms used for evaluation tests.

The test was divided into two parts; the first dealing with the ability to perceive room size when only utilizing real-time sound sources. For this part rooms with a wider spectrum of characteristics were used , the opera hall and the great hall with $RT_{60} \geq 1.4$ s and the office and lecture room with RT_{60}<0.5 s. The subjects were shown pictures of different rooms and had for each presented environment choose one of the depicted spaces which they sensed the closest to what they were hearing.

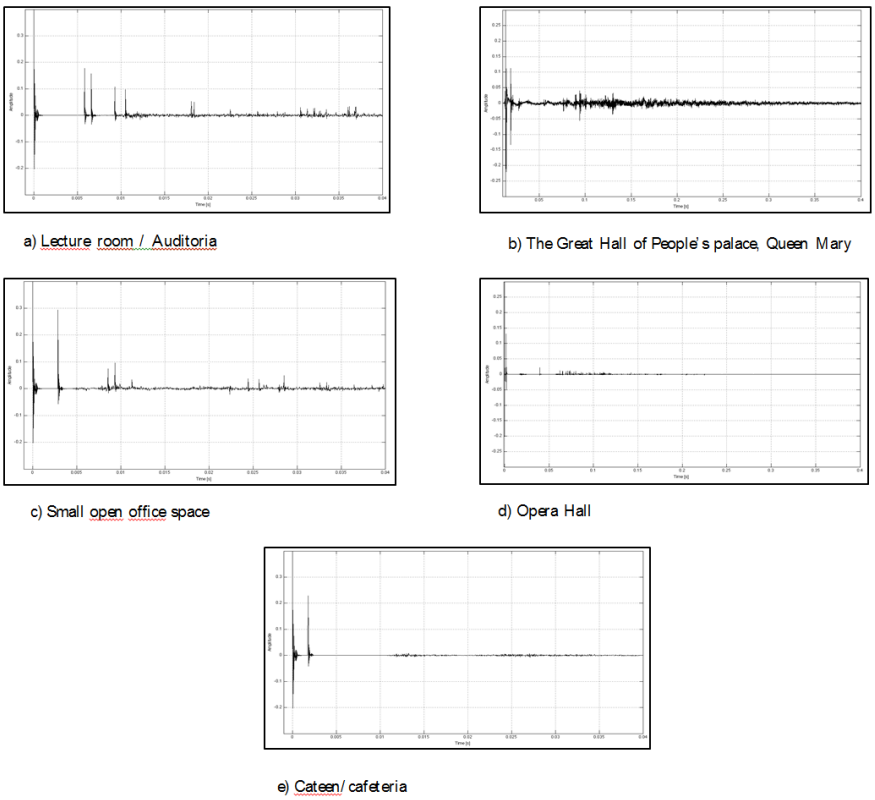

Figure 9.2: a)-e) Early parts of the impulse responses used for listening test including the direct sound for all cases. Al arel normalized with a direct sound equal to one except for the measured Great hall. Shown time lengths as well as amplitudes have been adjusted for visualization purposes.

The second part of the listening test was designed to evaluate the interactive element compared to using pre-recorded wav-files. Again four environments were used, the cafeteria, the smaller auditorium, the opera hall and the small open office space. Minimizing the time between presented spaces, the first section of testing was made using only pre-recorded sound sources switching to using real-time auralization in a latter section. When using the pre-convolved sound sources, the source position was determined arbitrarily in the environment. For each environment the user had to answer if the room volume was perceived as big or small and if the sounds heard within the room where perceived as either hard or soft. On a running scale from 1-5 they also had to judge the perceived naturalness of the environment, 1 being completely unnatural to 5, similar to a real environment experience. Again these factors can be hard to judge given the situation although for comparative purposes these where used but with some caution. Finally they had to correlate the perceived environment to different types of venues and rate the correspondence to their judgment of different environments from 1-5 in fixed steps. A number of different environments where written down and the subject had to place the simulated environment on the scale for each. As spaces used for different

purposes can vary vastly in size and geometry, indication where given as to what size the exampled room had, for example: a smaller lecture room or a larger exhibition hall. The results of the two different parts of the test were later compared. Once again dividing the different environments into groups of large, small and medium sized, judging the correlation between exampled rooms and the simulated one's.

9.3 Results of the evaluation test

9.3.1 Part 1 - Evaluating room size using only real-time synthesis

From the first section of the test where the subjects had to correlate depicted rooms with what they were hearing, it seemed easier to separate the larger rooms (the opera and great hall) from the medium-sized, and small sized rooms. Greater confusion seemed to occur when discriminating between the simulated small rooms with the depicted medium sized, where the answers had around a 50% spread. Results are presented in table 9.3.

9.3.2 Part 2 - Comparing real-time and pre-convolved sound sources

As can be seen from the results presented in table 9.4, the subjects had no problem distinguishing room size of the opera hall either with pre-convolved sound sources or with the real-time auralization.

Table 9.3: Test results part 1, perception of room size using only real-time auralization.

	RT_{60}	Approx. volume, $[m^3]$	Mean absorption*	Perceived Size Large	Medium	Small
Great Hall	2.1	≤ 15000	-	9	0	0
Opera Hall	1.4	15000	35%	7	0	2
Open-plan office	0.4	450	43%	0	4	5
Lecture room	0.4	400	35%	0	5	4

Table 9.4: Results from subjective listening test section 2, part 2.

	RT_{60}	Approx. volume, $[m^3]$	Mean absorption*	Perceived Size Pre-convolved sound source		Real-time sound source	
				Large	Small	Large	Small
Opera Hall	1.4	15000	35%	8	1	8	1
Canteen	0.4	3600	45%	4	5	3	6
Open-plan office	0.4	450	43%	1	8	1	8
Lecture room	0.4	400	35%	1	8	2	7

The canteen or cafeteria, although large in volume had unfortunately a very high mean absorption, resulting in the short reverberation time. The answered room size varied, more answered that it was smaller than large using real-time convolution. Tendencies where shown to judge the environments as softer using the real-time convolution, figure 9.3.

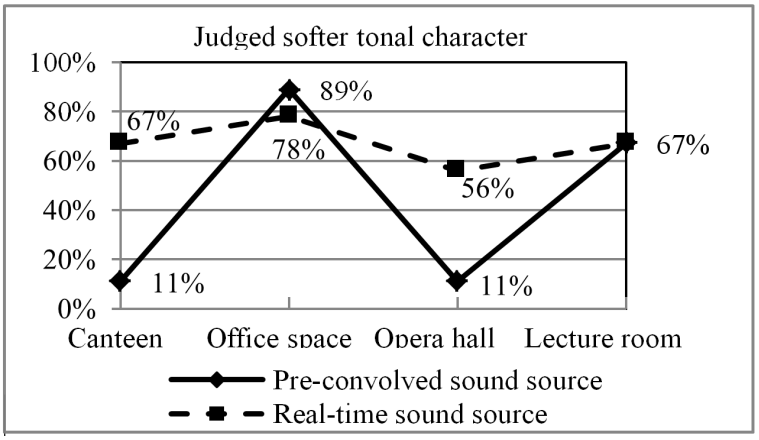

Figure 9.3: Judged tonal character, comparison between using the pre-convolved sound source and the real-time sound source. Results shown out of a 100% of the subjects i.e. 67% means amount of subjects judging it as soft whilst 33% judged it as hard.

In other results fewer differences were seen. Naturalness or authenticity ratings where overall deemed higher using the real-time convolution than when using pre-convolved ones. One can however not exclude that the subject had ideas about the aim of these tests, which might have had an effect on the results. When having to correlate the perceived room size to different exampled environments, differences in answers could be seen, however no distinct tendencies could be distinguished.

CHAPTER 10

Discussion

10.1 Design criteria for the auralization framework

One of the most important factors to consider when designing a real-time audio application for perception-based analysis is the time delay caused by signal processing. In this case the only time delay should be that of the initial time delay gap, determined by the RIR used. Acceptable added time delays should be below that of audibility. In this case, Haas criteria for audible echo is not sufficient as ' lagging' effects could be heard already at lesser time delays. Therefor the added time due to processing should preferably be less than 20-30 ms. If also removing the initial time delay gap from the RIR one can add to this time buffer. Adjustable block-size of the convolution algorithm can make auralization of smaller rooms more time efficient, reducing the block-size and the inherent time delay.

Another is to prepare the RIR used for simulation. As the subject will act as sound source and receiver the positioning of these should correspond to the distance of the subject's mouth and ear. Modelling the RIR this way creates problems concerning large dynamical differences between direct and reverberant sound. Approximations with this distance are deemed necessary and from user test results acceptable. The auralization framework is prepared to also use the Z-channel of the B-format to reproduce a three-dimensional sound field giving height information. At the time of writing, loudspeakers in the ceiling of the sound design lab are being installed making it possible to reproduce a hemisphere sound field in the future. Measurements suggests however that more feedback oscillations would then occur requiring a closer distance between the microphone and subject as well as additional filtering.

10.2 Reproduction level and the acoustic feedback problem

As the problem with acoustic feedback is dependent on needed amplification of the loudspeaker system, stated by the open-loop gain factor, $qH_{M(w)}$, this needs to be known. The needed amplification is in turn dependent on the SPL difference between direct sound and reflected sound, determined by the RIR used for simulation, and will thereby differ depending on simulated environment. If accepting the approximations relating to the added distance between source and receiver, and determining that this relationship should hold at the position of the talker as well, this will state which reproduction level is necessary for the loudspeaker system, and thus the needed amplification. Still this needed amplification is dependent on which RIR is used and how high the sound pressure level of the talker will be, making it difficult to predict to what extent acoustic feedback will occur and also to use above stated criterion and estimations. However, one should still strive to make the relationship between talker sound pressure level and that of the loudspeaker system as substantial as possible at the point of the microphone. From measurement results the microphone should have a narrow pick-up range and be placed as close to the talker as possible without making its presence to apparent. During measurements it was clear that multiple narrow band oscillations occurred when increasing amplification. This in turn implies that acoustic feedback control needs to be applied to ensure a stable system, handling different SPL differences of the RIR and source signal levels. As the problem seems to be that of acoustic feedback, not acoustic echo, a parametric-filter based equalizer can be used. Adaptive filtering, like the LMS method could also be used, but configuring the reference signal recording, used to produce the filter, is a tedious task when the signal needs to be recorded in the same enclosure as the wanted signal.

10.3 The subjective evaluation

Even though perceiving room size being one of the fundamental qualities of room perception, it is even for the acoustician hard to predict exact reverberation time, even dividing the different venues into large groups. Giving the subjects clues in form of depicted venues can be both helpful and detrimental for the evaluation. However, results from subjective evaluation test show differences in the perception of environments used for simulation differ using real-time sound sources and pre-convolved ones, suggesting further studies be made on this topic. A part of the easier recognition of large halls might possible be due to the deteriorated speaking conditions, another possible reason is also the vast difference between the listening rooms acoustical characteristics and the simulated one's. This however would imply that the subject has coupled reverberance with room size. As this was a fact using both pre-convolved sound sources and the real-time application, no discrepancy could be made between the methods. However it indicates that the perception of the sound field is not diminished by the added activity of the user. A possible reason for the small room confusion might be due to the listening

room and the simulated room having little difference in reverberation time. Subject's response to the tonal character of the environment, judging the environments in general as softer using the real-time auralization, might be due to the added control. It can however not be excluded that the characteristics of the sound source used for the pre-convolved auralization could have participated to these results. The authenticity ratings are hard to judge as different meanings could be put to the expression. The familiarity of one's voice could have a possible effect to the higher ratings when using real-time convolution. Clear indication of appreciation could although be confirmed as a large majority of the subjects responded they preferred hearing the simulated environments with the possibility to contribute to the environment, and being a part of it.

CHAPTER 11

Conclusion

11.1 the Auralization framework

An auralization method utilizing real-time sound sources has in this project been implemented and evaluated. The application, developed and compiled in the open source software Pd (Pure Data) is at this point able to auralize rooms with a room impulse response (RIR) length of up to 3 s, without any noticeable delays. At present, only static source and receiver positions are used, calculating and preparing the RIR offline. The application has been implemented in a controlled listening environment using an ordinary personal computer, a narrow-pick up microphone and a multi-channel loudspeaker system as well as feedback cancellation by a parametric-filter equalizer. Approximated RIRs have been necessary to use, calculating the RIR using a larger distance between source and receiver than that between the subject's mouth and ears (0.5 m instead of 0.1 m). From evaluation tests, a large majority of the subjects responded that they appreciated experiencing the simulated sound fields this way. This combined with high ratings of authenticity suggests that further investigations should be performed.

11.2 Choice of equipment and acoustic feedback problem

The microphone used should have a narrow shotgun-like directivity, and placed close to the talker (i.e. the subject). As shotgun-directivity microphones usually have both a distinct main-lobe and back-lobe at both ends of the microphone, it should be placed so that the main-lobe of the loudspeakers are in line with the microphone's sides. Amplification of the loudspeaker reproduction level is determined by the RIR and level of the sound source. As these varies for different environments, the amount of feedback is hard to determine although measurements show multiple feedback oscillations for most RIRs used. This requires measures of control in form of a parametric equalizer, ensuring stability by detecting oscillating frequencies rapidly.

11.3 Short notes on a transportable auralization set-up

As for compiling a transportable auralization set-up it can be a tedious task since many factors will affect the result. The room will still have to be sufficiently damped, with a lower RT_{60} than that of the modelled environment. Strong early reflections should be avoided and the background noise level should be kept low. As for reproduction system, possibly a stereo-channel loudspeaker system could be used, however this will not create the same feel of spaciousness as the multi-channel case. Usually these are implemented using cross-talk cancellation, if although the room gives rise to reflections the cross-talk cancellation will fail. For a transportable solution preferably headphone-reproduction should be used, implementing correct head related transfer function filtering. If the source signal is recorded at a close enough distance, there should be minimal room influence. Calibration of correct reproduction level will be difficult to achieve, and approximations might have to be made. If using a loudspeaker set-up possibly both loudspeaker and microphone should be directive, minimizing the effects of the environment. Measurements of the single-channel case, suggests that the loudspeaker should preferably be placed in front of the subject. The microphone used should have a narrow pick-up and should be placed at a 45 degree angle to the subject, at close distance.

CHAPTER 12

Future Work

A more extensive evaluation should be made as to what effects the mode of interaction interaction has on the subjective perception of the sound field. Further studies could be made as to what way the inherent speech alteration when subjected to different environments affect the room assessment. In this study only the real-time sound source was present in the simulated environment. Added environmental sounds as well as possible other sound sources might add to the experience, making the sound field more complete and natural. Using several modalities of interaction would also be of interest, using for example the real-time sound sources combined with possibility to move around in the environment. Since different interactive methods have shown positive results, combining several modalities of interaction might take us closer to the goal of reproducing realistic and immersive sound fields.

Bibliography

[1] M. Kleiner, B.-I. Dalenbäck, P. Svensson, Auralization-an overview, J. Audio Eng. Soc 41 (11) (1993) 861–875.
URL http://www.aes.org/e-lib/browse.cfm?elib=6976

[2] L. Savioja, J. Huopaniemi, T. Lokki, R. Väänänen, Creating interactive virtual acoustic environments, J. Audio Eng. Soc 47 (9) (1999) 675–705.
URL http://www.aes.org/e-lib/browse.cfm?elib=12095

[3] T. Lentz, D. Schröder, M. Vorländer, I. Assenmacher, Virtual reality system with integrated sound field simulation and reproduction, EURASIP J. Appl. Signal Process. 2007 (1) (2007) 187–187.
URL http://dx.doi.org/10.1155/2007/70540

[4] D. J. Furlong, M. P. Doyle, E. Kelly, C. J. MacCabe, R. MacLaverty, Interactive virtual acoustics synthesis system for architectural acoustics design, in: Audio Engineering Society Convention 93, 1992.
URL http://www.aes.org/e-lib/browse.cfm?elib=6691

[5] R. Appel, J. G. Beerends, On the quality of hearing one's own voice, J. Audio Eng. Soc 50 (4) (2002) 237–248.
URL http://www.aes.org/e-lib/browse.cfm?elib=11084

[6] K. Ueno, H. Tachibana, Experimental study on the evaluation of stage acoustics by musicians using a 6-channel sound simulation system, Acoustical Science and Technology 24 (3) (2003) 130–138.

[7] M. Kleiner, Acoustics and Audio Technology, Acoustics: Information and Communication, J. Ross Pub., 2011.

[8] B. Shinn-Cunningham, Acoustics and perception of sound in everyday environments, in: Proceedings of the 3rd International Workshop on Spatial Media, Aisu-Wakamatsu, Japan, 2003.

[9] T. van Watershoot, M. Monnen, Fifty years of acoustic feedback control: State of the art and future challenges, in: Proceedings of the IEEE, Vol. 93, 2011, pp. 288–327.
URL `http://dx.doi.org/10.1109/JPROC.2010.2090998`

[10] E. Hansler, G. Schmidt, Acoustic Echo and Noise Control: A Practical Approach, Wiley-Interscience, 2004.

[11] P. Svensson, On reverberation enhancement in auditoria (1994).

[12] H. Kuttruff, Acoustics: An Introduction.

[13] M. Frigo, S. G. Johnson, FFTW: An adaptive software architecture for the FFT, in: Proc. 1998 IEEE Intl. Conf. Acoustics Speech and Signal Processing, Vol. 3, IEEE, 1998, pp. 1381–1384.

[14] R. Stewart, M. Sandler, Database of omnidirectional and b-format impulse responses, in: Proc. of IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP 2010), Dallas, Texas, 2010.

APPENDIX A

Measurement Set-Up, Feedback Measurements in the Semi-Anechoic Chamber

A.1 Equipment data

Table A.1: Equipment used for acoustic feedback measurements - semianechoic chamber

Type	**Model**	**Manufacturer**
Talker source & RIR reproduction loudspeaker	Mask 6T-BL	Apart- Audio
Line loudspeaker, RIR reproduction loudspeaker	COLW-101	Apart- Audio
Loudspeaker amplification	PA2240BP	Apart-Audio
Microphones:		
Capsules &	CK31-33, CK47	AKG
Preamplifier	HM1000	AKG
Capsule &	Type 4189	Brüel & Kjaer
DSP	Soundweb BLU-101	BSS Audio
Measurement software	Room Capture	Wave capture
Data treatment	Matlab	Mathworks

A.2 Microphone directivity data

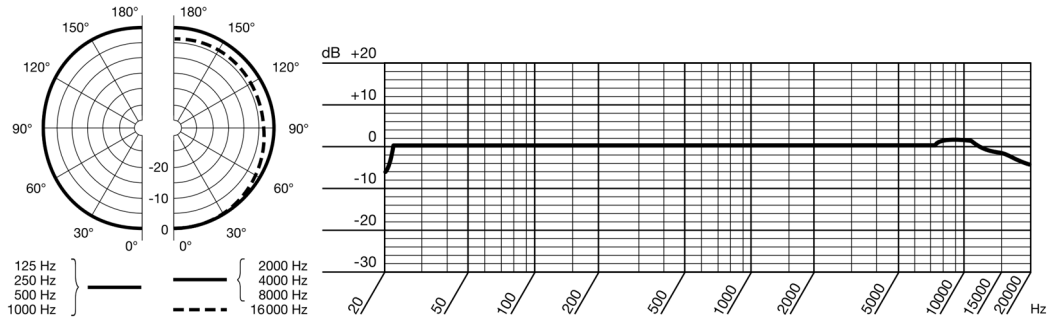

Figure A.1: Omni-directional capsule, CK 32

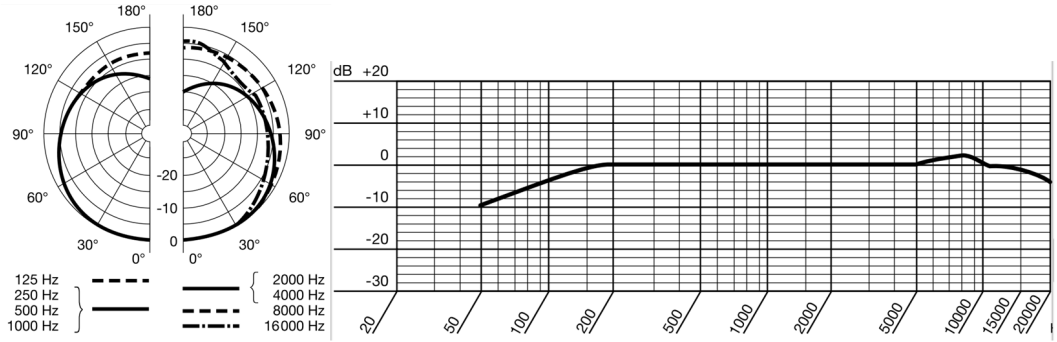

Figure A.2: Cardioid capsule, CK 31

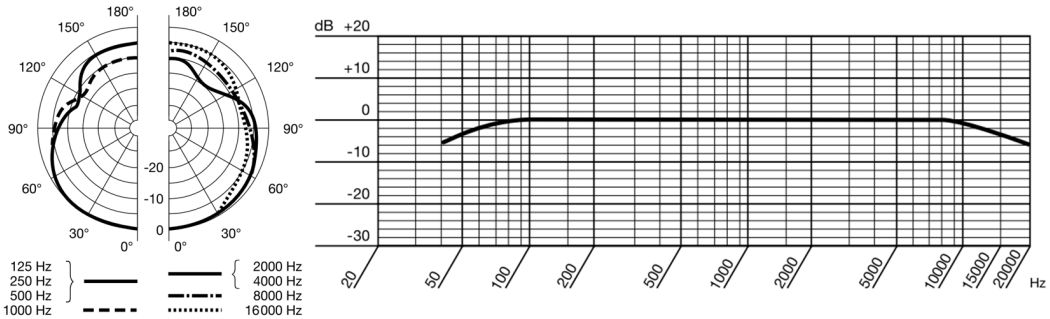

Figure A.3: Hyper-cardioid capsule, CK 33

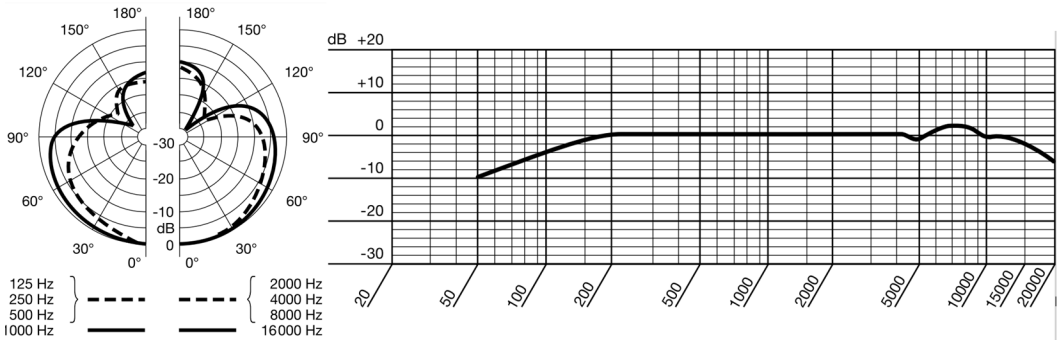

Figure A.4: Hyper-cardioid (shotgun) capsule, CK 47

A.3 Measurement set-up

Source signal: Sine Sweep
No. of averages: 4
Frequency range: 20-20000Hz
Time windowing: 200ms

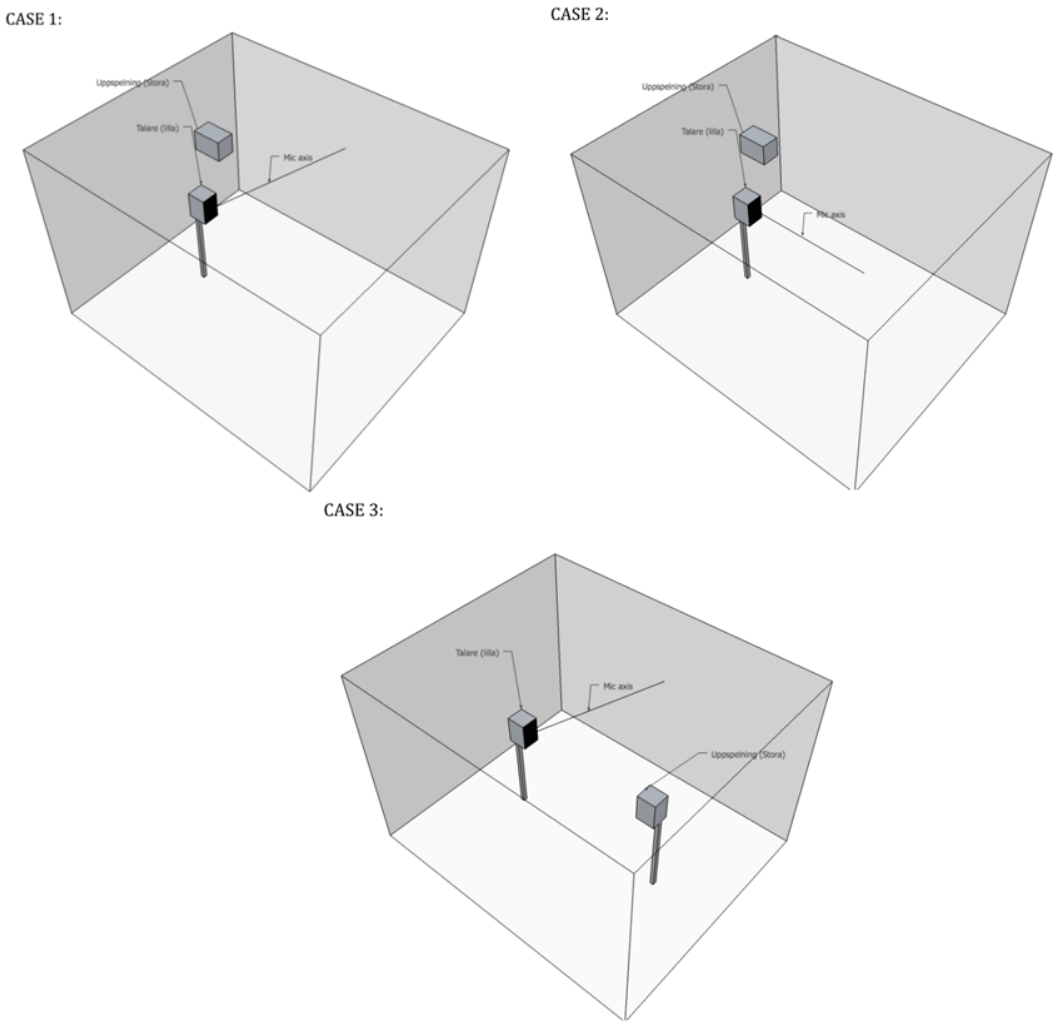

Figure A.5: Cases, measurement configuration

A.4 Measurement results

Exampled results from measurements

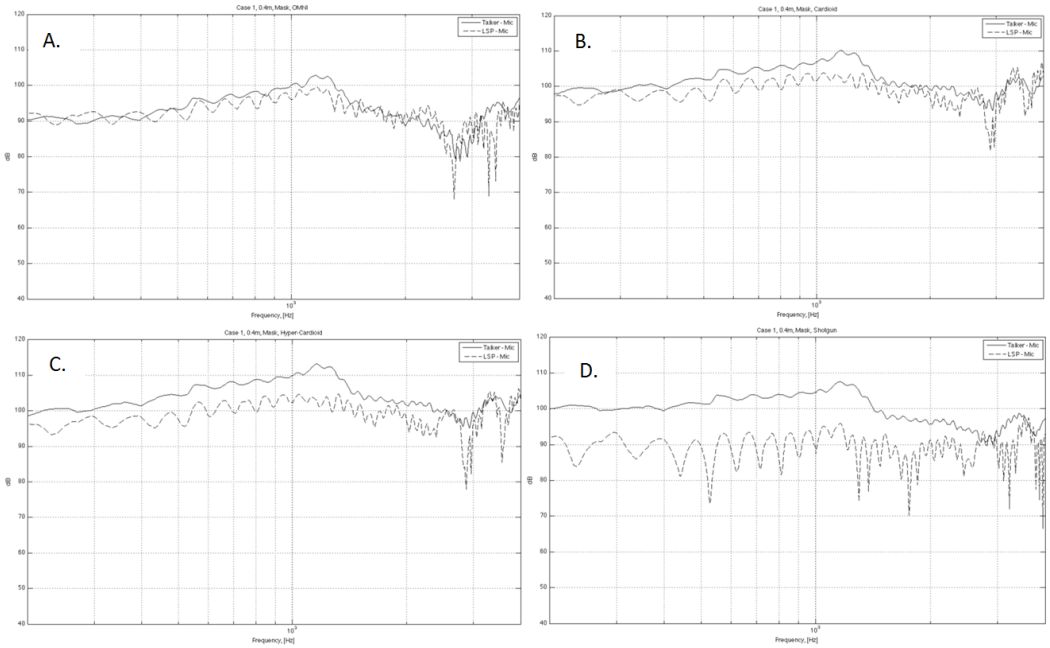

Figure A.6: Case 1: Transfer functions between talker source loudspeaker and microphone compared to transfer function between reproduction loudspeaker(omni-directional) and microphone. A, using the CK31 capsule, B, using the CK32 capsule, C, using the CK33 capsule and D, using the CK47 capsule Microphone is placed 0.4m from the talker source.

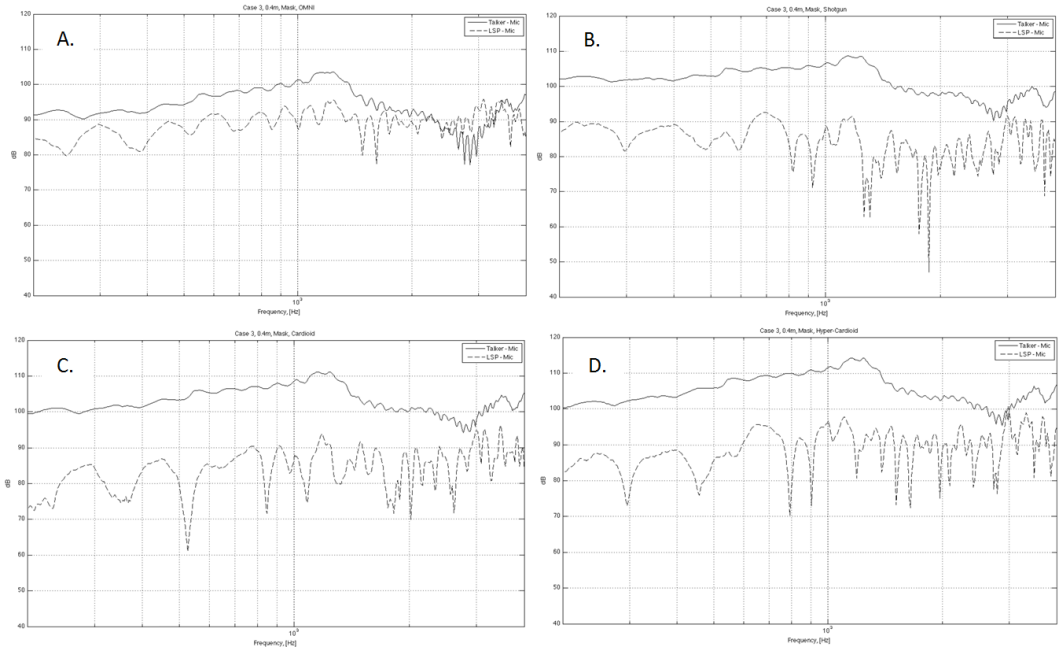

Figure A.7: Case 3: Transfer functions between talker source loudspeaker and microphone compared to transfer function between reproduction loudspeaker (omni-directional) and microphone. A, using the CK31 capsule, B, using the CK32 capsule, C, using the CK33 capsule and D, using the CK47 capsule Microphone is placed 0.4m from the talker source.

APPENDIX B

Measurement Set-Up, Feedback Measurements in the Sound Design Lab

B.1 Equipment data

Table B.1: Equipment used for acoustic feedback measurements -multichannel reproduction

Type	**Model**	**Manufacturer**
Simulated talker source	Mask 6T-BL	Apart- Audio
Loudspeaker amplification	PA2240BP	Apart-Audio
RIR Reproduction	Pre-installed multichannel system	Ino-Audio
Microphones:		
Capsules &	CK31-33, CK47	AKG
Preamplifier	HM1000	AKG
Capsule &	Type 4189	Brüel & Kjaer
Measurement software	Room Capture	Wave capture
Data treatment	Matlab	Mathworks

Figure B.1: The sound design lab at Konstfack.

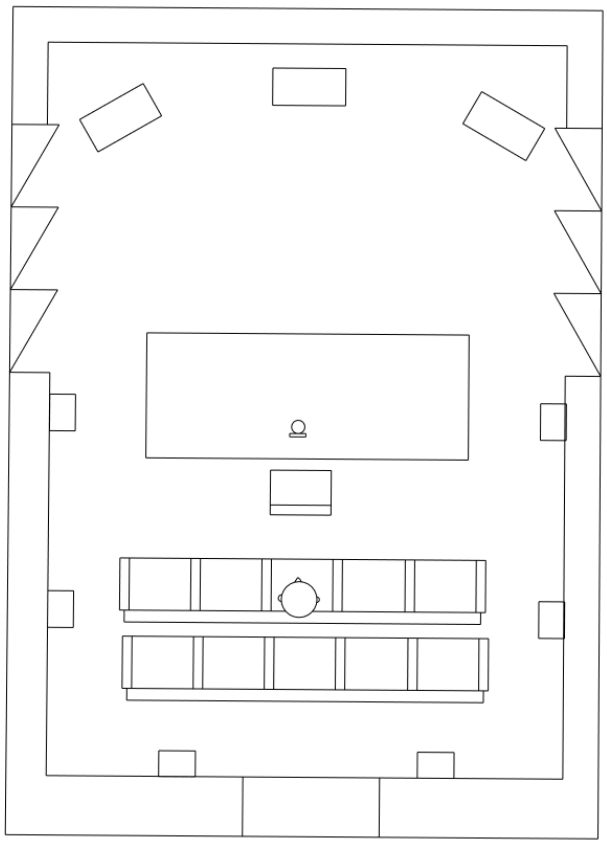

Figure B.2: Loudspeaker set up at the Sound Design Lab. Left surround and right surround signal are fed to clusters of three loudspeakers.

B.2 Measurement results Konstfack, multichannel loudspeaker set-up

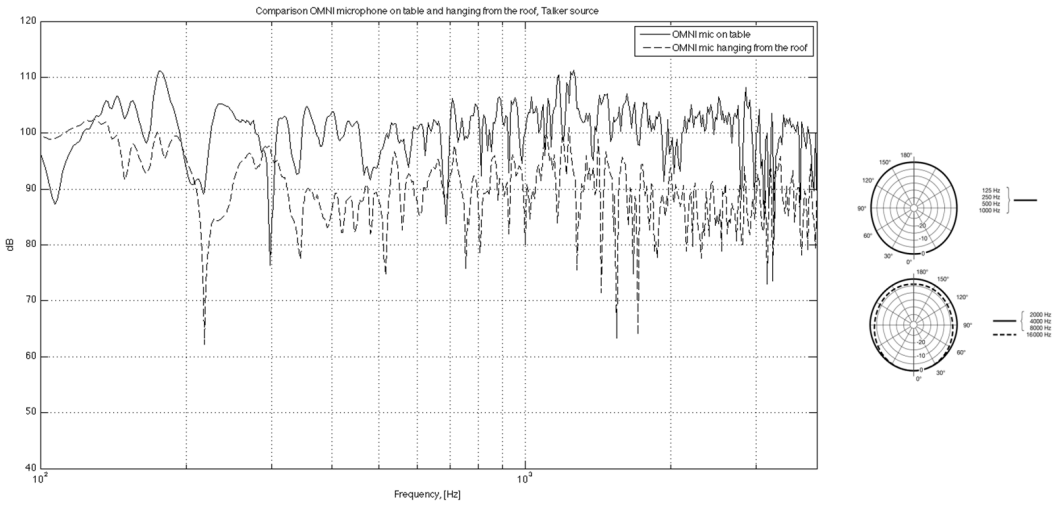

Figure B.3: Omni-directional capsule, CK 32, comparison hanging from the ceiling and placed on table

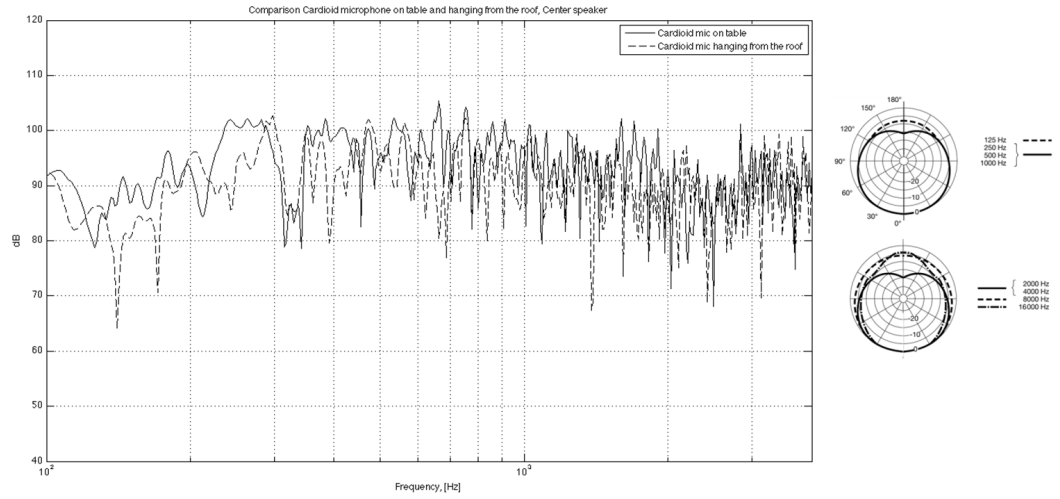

Figure B.4: Cardioid-directional capsule, CK 31, comparison hanging from the ceiling and placed on table

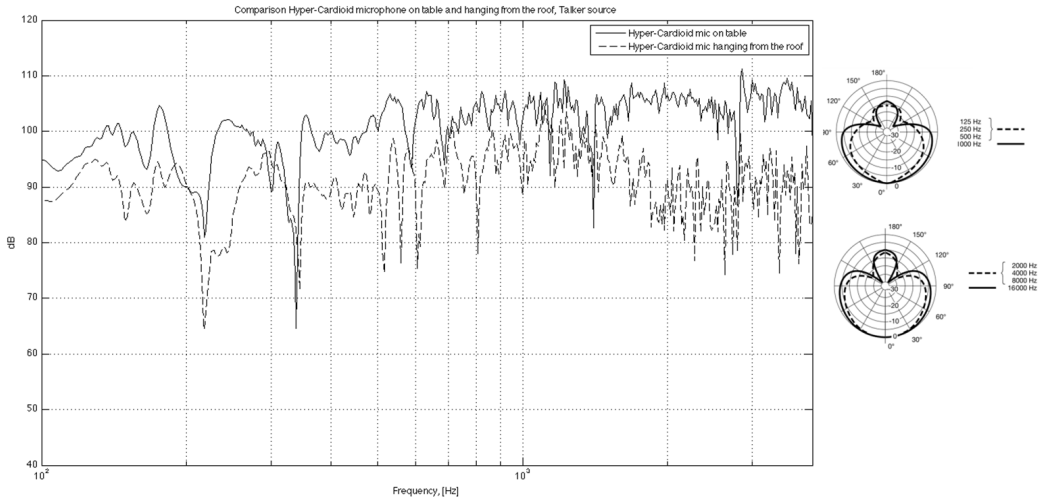

Figure B.5: Hyper-cardioid-directional capsule, CK 33, comparison hanging from the ceiling and placed on table

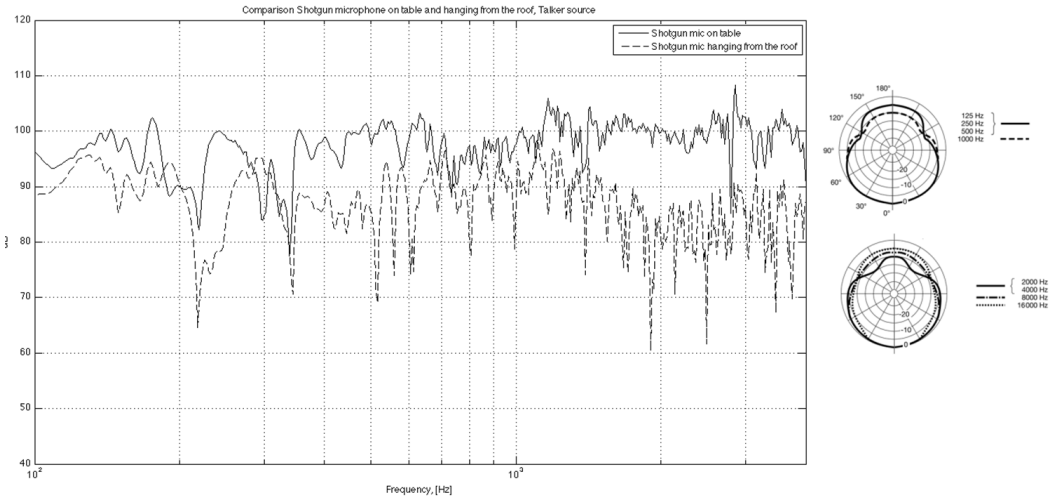

Figure B.6: Shotgun-directional capsule, CK 47, comparison hanging from the ceiling and placed on table

APPENDIX C

Listening Test Questionnaire

ENKÄT 2011-12/11 – 13/11

ENKÄT

1. Deltagarinfo

Prov nr:____ Datum:_____

Sysselsättningsområde:_____

Lider du av någon hörselproblematik? JA NEJ

Information:

I det här testet kommer du att få uppleva olika ljudmiljöer genom att enbart lyssna och ibland själv delta i miljön. För varje ljudmiljö ställs några kort frågor i denna enkät.

ENKÄT 2011-12/11 – 13/11

AVSNITT 1

Du kommer nu få lyssna på olika rum. Du kommer att kunna interagera med ljudmiljön genom att göra ljud, prata, klappa händerna med mera. Samtidigt visas olika rumstyper på bildskärmen. Ange vilken bild du anser bäst passar det upplevda rummet vad gäller upplevd storlek genom att ringa in respektive rumsnamn.

Observera nummerordningen!

Ljudmiljö 1

1	3	5
2	4	6

Ljudmiljö 3

1	3	5
2	4	6

Ljudmiljö 2

1	3	5
2	4	6

Ljudmiljö 4

1	3	5
2	4	6

ENKÄT 2011-12/11 – 13/11

Avsnitt 2

Detta avsnitt består av **två** olika delar, och du kommer att lyssna på 4 ljudmiljöer för respektive del.
I del 1 kommer du enbart lyssna, i del 2 gör du ljud så som att klappa händerna, prata eller dylikt.

Punkt 1-2: För varje ljudmiljö ska nedanstående frågor besvaras (obs, en sida per ljudmiljö). Välj för varje kategori den egenskap hos ljudmiljön som du anser bäst överensstämmer med det du upplever.

Punkt 3: Rangordna på skalan 1-5 (kryssa i rutan) hur pass likt din upplevelse av rummet

Exempel:

1. Ange vilken egenskap som bäst stämmer överens med din upplevelse med ett kryss:

	Litet	Stort
Rummets volym:	☐	☐

	Hårt	Mjukt
Rummets klang	☐	☐

2. Hur naturlig (autentisk) upplever du miljön?

Onaturlig — Naturlig

1 2 3 4 5

3. På skala nedan, rangordna i vilken mån upplevelsen av den uppspelade ljudmiljön stämmer överens med respektive rums typ:

	Inte alls	Stämmer sämre	Varken eller	Stämmer	Stämmer mycket bra
Restaurang/Matsal	☐	☐	☐	☐	☐
Aula/hörsal	☐	☐	☐	☐	☐
Konserthall/opera hall	☐	☐	☐	☐	☐
Ljusgård/öppen yta	☐	☐	☐	☐	☐
Kontorslandskap	☐	☐	☐	☐	☐

APPENDIX D

Auralization Set-Up

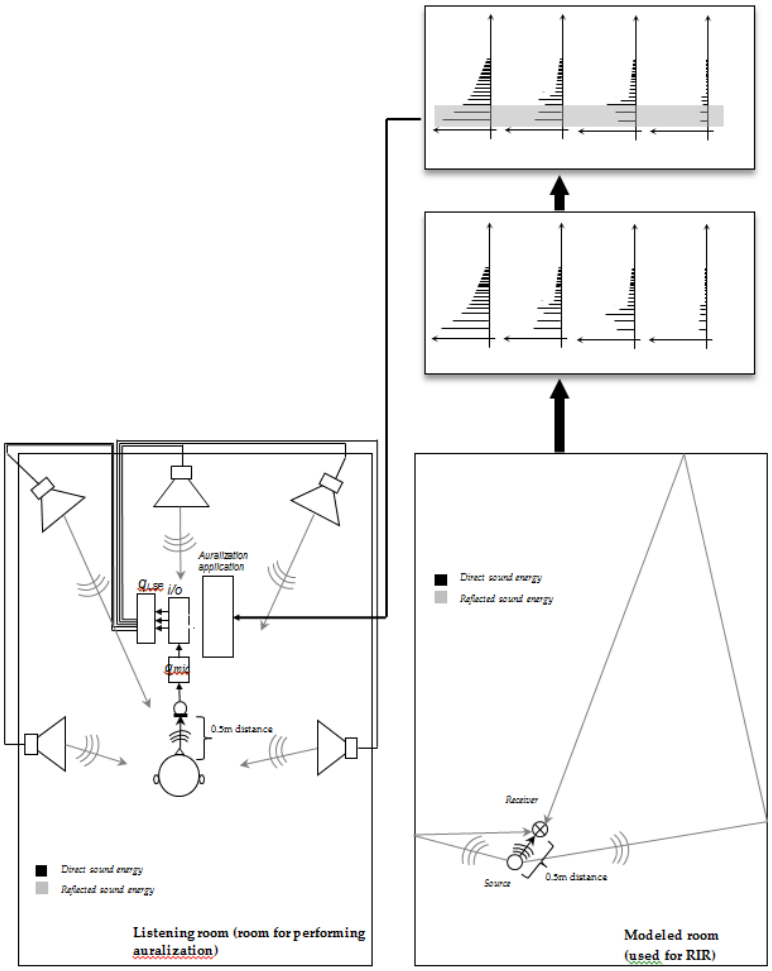

Figure D.1: Schematic of the auralization set-up