LUND UNIVERSITY

A COMPARISON OF BINAURAL RECORDINGS PERFORMED IN SIX DIFFERENT ROOM CONFIGURATIONS

DENICE PERKHED

Engineering Acoustics

Master's Dissertation

DEPARTMENT OF CONSTRUCTION SCIENCES
DIVISION OF ENGINEERING ACOUSTICS

ISRN LUTVDG/TVBA--20/5059--SE (1-67) | ISSN 0281-8477
MASTER'S DISSERTATION

A COMPARISON OF BINAURAL RECORDINGS PERFORMED IN SIX DIFFERENT ROOM CONFIGURATIONS

DENICE PERKHED

Supervisor: Dr **ERLING NILSSON**, Div. of Engineering Acoustics, LTH, Lund.
Assistant Supervisor: Dr **NIKOLAOS VARDAXIS**, Div. of Engineering Acoustics, LTH, Lund.
Examiner: **DELPHINE BARD**, Associate Professor, Div. of Engineering Acoustics, LTH, Lund.

Printed by V-husets tryckeri LTH, Lund, Sweden, February 2020 *(Pl)*.

For information, address:
Division of Engineering Acoustics,
Faculty of Engineering LTH, Lund University, Box 118, SE-221 00 Lund, Sweden.
Homepage: www.akustik.lth.se

Abstract

Children in Swedish preschools are exposed to high noise levels that can cause hearing damage or voice related problems among both teachers and students, this since they have to raise their voice in order to make themselves heard in a noisy environment. Because of this it is of interest to be able to predict the sound environment in preschools. A way to do this is through convolving binaural recordings of typical preschool noise with simulated binaural impulse responses of a preschool's premises. In order for this to work the binaural recordings has to be performed in an anechoic chamber. However, it is not always considered safe to perform binaural recordings of preschool children in an anechoic environment since the floor often consists of a net. A semi-anechoic environment, however, has a real floor and could possibly be created on a preschools own premises by applying sound absorbing elements on the walls. But will binaural recordings performed in a semi-anechoic environment give the same results as binaural recordings performed in an anechoic environment after convolution? It is assumed that if no difference is perceived between the binaural recordings performed in the two different environments there should not be a perceived difference between them after convolution as well.

In order to investigate this, binaural recordings were performed in six different room configurations with different absorption, from fully reverberant to anechoic. The recordings were performed by using an artificial head. Room acoustic parameters, such as reverberation time, speech clarity and sound pressure level was measured in all room configurations at the same places as the artificial head and the loudspeaker had been placed. The results of the measurements showed that the reverberation time was the shortest in room configuration C, which represented a semi-anechoic chamber. In this room configuration there was no values for the speech clarity for frequencies above 1000 Hz, this suggests that there is a lack of late reflections.

A listening test was then performed, all room configurations were compared to the anechoic recording in order to investigate if there was a perceived difference and how large that difference was in that case. The results of the listening test were then analyzed statistically through the method of t-test. Results showed that there was a significant difference between binaural recordings performed in all of the six room configurations and binaural recordings performed in an anechoic chamber.

To characterize the absorption in each room an average absorption coefficient was calculated with the help of Eyring's formula. The calculations showed that the value of the average absorption coefficient was largest for room configuration C where it had a value of approximately 0,7. In anechoic chambers the average absorption coefficient usually has a value around 1.

The significant difference between how binaural recordings performed in room configuration C and binaural recordings performed in an anechoic chamber are perceived suggests that it is

the early reflections that are of importance for the perceived difference. This because there is a lack of late reflections in this room configuration for frequencies above 1000 Hz.

Sammanfattning

Barn i svenska förskolor utsätts för höga ljudnivåer som kan orsaka hörselskador och röstrelaterade problem bland både lärare och elever. Detta eftersom de behöver höja rösten för att göra sig hörda i en bullrig miljö. På grund av detta är det av intresse att kunna förutse ljudmiljön i förskolor, ett sätt att göra detta är genom att falta binaurala inspelningar av typiska förskoleljud med simulerade binaurala impulssvar från en förskolas lokaler. För att det här ska fungera behöver dock de binaurala inspelningarna vara utförda i ett ekofritt rum. Det anses dock inte alltid vara säkert att genomföra binaurala inspelningar av förskolebarn i ett ekofritt rum eftersom golvet här ofta består av ett nät. Ett semiekofritt rum har emellertid ett riktigt golv och kan möjligtvis skapas i en förskolas egna lokaler genom att applicera ljudabsorberande element på väggarna. Men kommer binaurala inspelningar som utförs i en semi-ekofri miljö ge samma resultat som binaurala inspelningar som utförts i en ekofri miljö efter att de faltats? Det antas att om ingen skillnad uppfattas mellan de binaurala inspelningarna som utförs i de två olika miljöerna, så finns det inte heller en upplevd skillnad mellan dem efter faltning.

För att undersöka detta utfördes binaurala inspelningar i sex olika rumskonfigurationer med olika absorption. Inspelningarna utfördes med hjälp av ett konsthuvud. Rumsakustiska parametrar så som efterklangstid, taltydlighet och ljudtrycksnivå mättes i alla rumskonfigurationer på samma positioner där högtalaren och konsthuvudet varit placerade. Mätningarna resulterade i att rumskonfiguration C, som representerade ett semi-ekofritt rum, hade lägst efterklangstid av samtliga rumskonfigurationer. I detta rum saknades även värden för taltydligheten för frekvenser över 1000 Hz. Detta tyder på en avsaknad av sena reflexer i rummet.

Sedan utfördes ett lyssningstest där binaurala inspelningar från alla rumskonfigurationer jämfördes med binaurala inspelningar utförda i ett ekofritt rum. Detta gjordes för att undersöka om det fanns en upplevd skillnad mellan inspelningar utförda i de olika rumskonfigurationerna och ekofria inspelningar och hur stor denna skillnad i så fall var. Resultatet av lyssningstestet analyserades sedan statistiskt genom ett t-test. Resultaten av t-testet visar att finns en signifikant skillnad mellan binaurala inspelningar utförda i samtliga sex rumskonfigurationer och binaurala inspelningar utförda i ett ekofritt rum.

För att karaktärisera absorptionen i varje rumskonfiguration beräknades en medelabsorptionskoefficient med hjälp av Eyrings formel. Beräkningarna visade att medelabsorptionskoefficienten var högst för rumskonfiguration C där den hade ett värde av 0,7. I ekofria rum brukar detta värde ligga runt 1.

Att det är en signfikant skillnad mellan hur binaurala inspelningar utförda i rumskonfiguration C och binaurala inspelningar utförda i ett ekofritt rum upplevs tyder det på att det är de tidiga reflexerna som är viktiga för den upplevda skillnaden. Detta eftersom det finns en avsaknad av sena reflexer i rumskonfiguration C för frekvenser över 1000 Hz.

Acknowledgments

This Master thesis is the final part of the civil engineering program at Lund university and includes 30 credits. The work has been performed within the division of Engineering Acoustics at Lund university and in collaboration with ÅF Sound and Vibrations during the spring and summer of 2019.

During the work with the thesis all support and shared knowledge from both Lund university and ÅF sound and vibrations has been very much appreciated. I would like to thank my supervisor Erling Nilsson for all the support, guidance and feedback throughout the work with the thesis. Thanks to Nikolas Vardaxis for taking the time to give feedback and discuss the work.

I would also like to express my gratitude to all personnel at ÅF Sound and Vibrations in Malmö for all their advices and support during the thesis.

Furthermore, I would like to thank everyone that took the time to participate in the listening test and answering the questionnaires that came with it.

Finally, I want to thank my family and friends for their amazing support throughout the entire study period. Without you this would not have been possible.

Lund, 9 September 2019

Denice Perkhed

Table of content

1. Introduction

1.1. Background

Binaural hearing is the ability to hear with two ears. The difference between the input signals between the two ears provides information that makes it possible to localize sound sources (Blauert, 2013). The idea behind binaural recordings is to record these signals and reproduce them exactly as they were (Hammershøi & Møller, 1992). It is assumed that with this technique the complete auditory experience could be reproduced, including timbre and spatial cues. Binaural recordings are usually obtained by recording with an artificial head with two microphones placed either at the entrances of the ear canals or at an arbitrary point along the ear canals (Xie, 2013).

Binaural recordings could be used for predicting the sound environment of rooms or buildings that have not yet been built. If the binaural impulse responses of the premises can be simulated, then they can be convolved with binaural recordings performed in an anechoic environment. By convolving these two it is possible to predict how the anechoic recordings will sound in the premises. Convolution practically adds the influence of the room to the anechoic recordings. However, to be able to do the convolution it is absolutely necessary to use binaural recordings performed in an anechoic chamber to eliminate the influence of the room on the recordings.

Binaural recordings take into account the filtering effect that the human anatomy has on sound, this is called the head related transfer functions (HRTFs) (Xie, 2013). Since the anatomy of each individual is unique it varies between individuals, it also varies considerably with growth (Fels, 2008). To measure the HRFT is the most common way to obtain individual HRTFs. The subject has to be sitting in an anechoic environment, microphones are then placed in the entrances of the ear canals. The position of the head has to be fixed during the measurement. Unlike adults, children are not able to participate in this kind of measurements. This is because of the environment in the anechoic chamber and the fact that the subject has to remain motionless during the entire measurement.

It is of interest to be able to predict the sound environment of preschools since the noise level in some Swedish preschools are high enough to be able to cause hearing damage (Sjödin, 2012). According to measurements performed by (Waye, Agge, Lindström & Hult, 2011) children in preschools are in average exposed to an equivalent sound level of 85 dBA and a level of maximum sound pressure levels of 117-118 dBA. A high sound level in classrooms and preschools can cause voice related problems among both teachers and students since they have to raise their voice in order to make themselves heard (Socialstyrelsen. 2008). A high level of background noise therefore often leads to increased sound levels because the people present in the room have to raise their voices to make themselves heard in a noisy environment.

A very noticeable effect of noise is that it makes it more difficult to perceive speech (Sjödin, 2012). It is usually said that the noise masks the wanted sounds. Children are sensitive to noise since they are developing their understanding of speech and their vocabulary (Socialstyrelsen. 2008). This makes it hard for them to fill in the parts of the speech that is masked by noise. Children also have a greater risk for hearing damage (Socialstyrelsen, 2008). A reason for this could be that children have a shorter ear canal compared to adults, the sound will therefore not be as muted when it reaches the eardrum as it will be among adults.

To be able to predict the sound environment of preschools anechoic binaural recordings of preschool noise are needed. However, it is not always possible to perform binaural recordings in an anechoic environment. Anechoic chambers can be hard to get a hold of or sometimes considered too unsafe. It is for instance not considered safe to perform binaural recordings of preschool children in an anechoic chamber, this since the floor often consists of a net and the children could perceive the anechoic environment as scary. In this case there is a risk that the children would not act natural and the recorded sound would probably not represent the sounds that would occur in a real preschool environment. A semi-anechoic chamber, however, has a real floor and could possibly be created on a preschool's own premises by applying sound absorbing elements on the walls and/or floor of a room. By creating a semi-anechoic environment in a room on a preschool's own premises, and perform the binaural recordings there, the children would probably feel safer and act more natural since they would be in an environment that they are familiar with and will have access to their usual toys. This could probably lead to more accurate results.

1.2. Problem formulation

The purpose of this thesis is to investigate how much absorption that has to be present in a room before no difference is perceived between binaural recordings performed in the room and binaural recordings performed in an anechoic environment. The absorption in a room is characterized by the average absorption coefficient, α_m. It can be calculated by rewriting Eyring's formula (Kuttruff, 1991).

$$T = -0{,}16 \cdot \frac{V}{S_t \cdot \ln(1-\alpha_m)} \rightarrow \alpha_m = 1 - e^{\frac{-0{,}16 \cdot V}{S_t \cdot T}} \qquad (1.1)$$

where
V is the volume of the room
T is the reverberation time
S_t is the total surface of the room boundaries (walls, floor and ceiling)

The result will show if a semi-anechoic environment is sufficient for performing binaural recordings and if this is a suitable method for investigating for instance, preschool noise.

How large does the absorption factor, α_m, of a room need to be before no difference is perceived between binaural recordings performed in the room and in an anechoic chamber?

1.3. Aims and objectives

The aim of the study is to investigate the possibility of using a semi-anechoic environment for performing binaural recordings of preschool noise. This will be done by investigating how large the absorption coefficient, α_m, must be in a room before no difference is perceived between binaural recordings performed in the room and recordings performed in an anechoic environment. Depending on how high the absorption needs to be, a conclusion can be made whether this is a good method for investigating preschool noise. If a too large amount of absorbers are needed it can be hard to create the semi-anechoic environment on the preschool's own premises and the method can thus be assumed to be inappropriate. It is assumed that if no difference is perceived between the binaural recordings performed in the two different environments there should not be a perceived difference between them after convolution with identical binaural impulse responses as well.

1.4. Limitations

The focus of the thesis is to investigate how large the absorption coefficient, α_m, must be before the sound environment in a room is perceived exactly as the sound environment in an anechoic chamber. The thesis will include a listening test and relevant statistical theory to give an answer to the issue.

The thesis will not focus on headphone-to-ear-canal transfer functions and how to compensate for them. There will be a theoretical study of the headphone equalization process, but it will not be done practically during the recording and playback process.

Since the recordings are performed in all the different room configurations there is no need for a convolution to be done. Therefore, convolution will not be included in this thesis.

1.5. Methodology

In order to investigate how large the absorption coefficient, α_m, must be in a room before binaural recordings performed in the room are perceived in the same way as binaural recordings performed in an anechoic environment, a literature study, binaural recordings and a listening test will be performed. The Binaural recordings are performed in six different room configurations from fully reverberant to fully anechoic, this will be done by varying the amount of absorbers present in a room from none at all to a lot. In each room configuration the room impulse response will be measured, from this, important acoustic parameters such as reverberation time, speech clarity and sound strength will be determined. The acoustic parameters will then be compared between the different room configurations. A listening test will be performed to determine how large the absorption coefficient needs to be before the binaural recordings performed in the room are perceived exactly as the anechoic binaural recordings, the listening test will be performed by the binaural recordings being played back through headphones. The listening test will consist of a group of people listening to the binaural recordings performed in the anechoic chamber compared to all the other room configurations, they will then answer a questionnaire. To evaluate the results of the questionnaire statistical theory must be studied. The method of paired comparison will be used, after that a t-value will be calculated to determine if there is a significant difference between binaural recordings performed in each of the six room configurations and binaural recordings performed in an anechoic chamber.

2. Theory

It is important to first have a basic understanding of how sound works and how it behaves in rooms before discussing binaural recordings. Here the basic characteristics of sound and some quantities commonly used in room acoustics will be introduced first. This is followed by a description of the functions of the human auditory system. Since the final consumer of any acoustic environment is always the listener it is also important to have a basic understanding of psychoacoustics and commonly used subjective quantities.

2.1. Fundamental acoustics

In this chapter, basic acoustic theory will be explained to give a better understanding of how sound works and behaves. Further down some fundamental theory of sound is introduced, including some common quantities used in room acoustics and what they depend on.

2.1.1. Sound

Sound waves

Sound is pressure variations in the air that usually is created by a vibrating object, these pressure variations in the air are perceived as sound by the human ear (Ginn, 1978). The vibrations give rise to sound waves that can travel through a medium. In room acoustics air is the only medium concerned (Kuttruff, 1991). The sound waves propagate through the medium by displacing particles. The displaced particle can then transfer momentum to a nearby particle and hence, a wave is created (Ginn, 1978). The sound wave then propagates outward from the vibrating object, it generally gets weaker the longer away from the source it travels (Moore, 2004). The molecules in the medium does not advance with the sound wave, they oscillate around a position of equilibrium. Since sound often is affected by reflection and diffraction caused by obstacles in its way, the sound that reaches the ear differs from the sound that was generated at the source. Any complex sound field can be considered as a superposition of many sound waves (Kuttruff, 1991).

Frequency

The rate of oscillation of the vibrating object determines the frequency of the created sound wave (Ginn, 1978). When the oscillation repeats itself, the motion is said to have completed one cycle. The number of completed cycles during one second is referred to as the frequency, *f*, of the sound and is given in the unit Hertz (Hz).

The period, *T*, is defined as the time it takes for the oscillation to complete one cycle. The relation between the frequency and the period can be described by the equation below:

$$f = \frac{1}{T} \tag{2.1}$$

Wavelength

The wavelength, λ, of the sound is defined as the length of a sound wave, it corresponds to the period time), T (Nilsson, Johansson, Brunskog, Sjökvist & Holmberg, 2008). The relationship between the wavelength and the frequency is as follows:

$$\lambda = \frac{c}{f}\ [m] \tag{2.2}$$

where
c is the speed of propagation.

As the equation shows, the wavelength is shorter for high frequencies and longer for low ones. The audible range between 20 Hz and 20 kHz corresponds to wavelengths between 17 meters and 17 millimeters (Åkerlöf, 2001).

Diffraction

When the sound waves encounter an obstacle that is much larger than the wavelength of the sound, the obstacle will have a shadowing effect and the sound waves will not reach all the way behind the obstacle (Nilsson et al, 2008). The shadowing effect will only be significant if the dimensions of the obstacle are very large in relation to the wavelength. Small objects will therefore practically not give rise to a shadowing effect. This occurs because of a phenomenon called diffraction; the direction of propagation will change when a sound wave encounters an obstacle. Thus, sound waves can travel around corners.

Decibel

Sound pressure, sound intensity and sound power are for convenience expressed by Pa, W/m^2 and W respectively (Ginn, 1978). To include the wide range of audible intensities a logarithmic scale is often used, the most common logarithmic scale for describing sound levels is the decibel scale. Decibel is a relative measurement, all measured sound pressures are expressed as a ratio relative to a reference pressure.

2.1.2. Reverberation time

Reverberation in a room occurs when sound is reflected numerous times at the room boundaries such as walls, floor and ceiling (Vorländer, 2008). Each time a sound wave is reflected by a surface a part of its energy is lost due to absorption, the sound wave will continue to be reflected by surfaces until it is gone (Kuttruff, 1991). The time it takes for the sound level to decrease 60 dB from the time that the sound source is turned off is defined as the reverberation time (Nilsson et al, 2008). The reverberation time depends on the volume of the room and the amount of absorbers present in the room. The more absorbers it is in the room the shorter the reverberation time will be. In rooms intended for children, elderly and people with impaired hearing a reverberation time below 0,5 seconds could be required (Arbetsmiljöverket, 2005). A longer reverberation time will lead to an increased sound pressure in the room, by shortening the reverberation time the risk of disturbances and hearing damage can therefore be decreased (Socialstyrelsen, 2008). A long reverberation time will also make it more difficult to perceive speech, this because a longer reverberation time means that it will take a longer time for speech and other sounds to decrease. This can result in masking of wanted sounds (Socialstyrelsen, 2008). The reverberation time can be calculated with Sabine's formula (Nilsson et al, 2008):

$$T = 0{,}16 \cdot \frac{V}{\alpha \cdot S} \tag{2.3}$$

where
V is the volume of the room
α is the absorption coefficient of the absorbing material
S is the surface of the absorbing material

or with Eyring's formula:

$$T = -0{,}16 \cdot \frac{V}{S_t \cdot \ln(1 - \alpha_m)} \tag{1.1}$$

where
V is the volume of the room
T is the reverberation time
S_t is the total surface of the room boundaries (walls, floor and ceiling)

2.1.3. Sound absorbers (Absorption coefficient)

A room is made up of several boundaries, such as walls, floor and ceiling, with different qualities. When sound waves encounter these boundaries a part of the sound energy is reflected back into the room while a part of it is absorbed either by being converted to heat or by being transmitted to the other side of for instance a wall, see Figure 2.1 (Kuttruff, 1991).

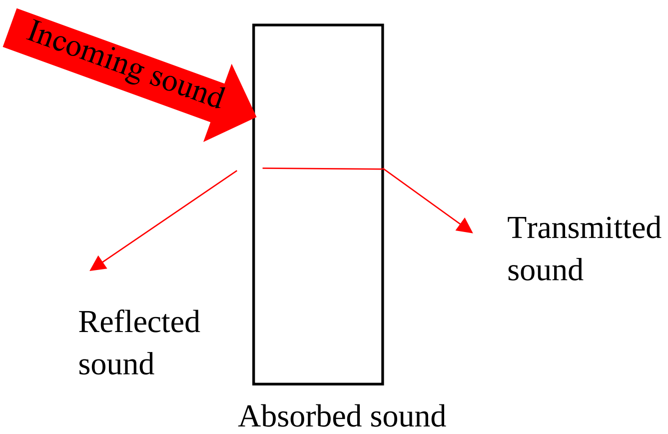


Figure 2.1. The soundwave after it encounters a room boundary.

All surfaces in a room absorb sound to some extent (Ginn, 1978). How much sound a surface absorb depends on what material it is made of, hard smooth surfaces will absorb less sound while soft and porous surfaces, like carpets or curtains, will absorb more sound. For a porous material to be able to absorb sound there must be open pores in the material (Åkerlöf, 2001). If the porosity of a material is too low the sound will be prevented from entering the material and will therefore be reflected back into the room instead. This could occur if for instance a porous absorbent is painted. If the material has a porosity that is too high the sound will go through the material without any friction occurring between the soundwaves and the material. If there is no friction between the material and the soundwaves there will be no sound absorption. When a sound wave hits this type of material it causes the air in the pores to vibrate, the flow resistance of the material limits the movement of the air particles. (Ginn, 1978). A part of the sound energy is thus converted to heat. Porous sound absorbing materials could be made of for instance mineral wool.

The absorption coefficient, α, of a material indicates how much of the incident sound that is absorbed by the material (Nilsson et al, 2008). The value of the absorption coefficient varies between zero and one for different materials, zero means no absorption while one means total absorption (Bernström, 1987). An open window is considered to be a perfect absorbent, $\alpha=1$, since the sound that hits the window will continue out of it, nothing will be reflected back into the room (Åkerlöf, 2001).The absorption coefficient for a certain material will vary with frequency (Meyer, 2009). This means that a material's sound absorbing ability also varies with frequency.

The equivalent absorption area, *A*, of a room describes the sound absorbing ability of a material of a specific size. (Meyer, 2009). It is calculated as below:

$$A = \alpha \cdot S \qquad (2.4)$$

where
S is the size of the surface that has the absorption coefficient α.

The total absorption of a room is the sum of the equivalent absorption area of all absorbents in the room, see equation below (Bernström, 1987). Here the absorption of all materials in the room is included, such as walls, ceiling, floor, windows, etc.

$$A_{tot} = \alpha_1 \cdot S_1 + \alpha_2 \cdot S_2 + \cdots + \alpha_n S_n \qquad (2.5)$$

Since the equivalent absorption area depends on the absorption coefficient and the size of the surface of the material, a large surface of a material with a low absorption coefficient could still have a large impact on the total absorption of the room.

2.1.4. Flutter echo

Flutter echo is a phenomenon that occurs when a room has two hard parallel walls while the rest of the room surfaces are either absorbing or diffusing the sound (Nilsson et al, 2008). This will result in the sound energy staying in the room in the form of reflections between the two walls. This can be avoided by diffusing or absorbing the reflections from the walls or by angle the walls.

2.1.5. Diffuse sound field

In a diffuse sound field, the amplitudes of the incident sound waves are uniformly distributed over all possible directions of incidence and the phases of the waves are randomly distributed (Kuttruff, 1991). A diffuse sound field is an ideal condition, it can be realized fairly well in some types of rooms, for instance reverberation chambers.

2.1.6. Free field

A free field is an environment in which there is no reflective surfaces, the sound will therefore not be affected by reflections. This can for instance be in an anechoic chamber (Kuttruff, 1991).

2.1.7. Near field

The sound field in a room can be divided into two parts, the sound waves that has just left the sound source and have not yet been reflected by the room boundaries and the sound waves that have been reflected one or more times (Bodén, Carlsson, Glav, Wallin & Åbom, 2012). The immediate vicinity of the sound source is known as the near field, this is the part of the sound field that is still unaffected by the room properties. It can be considered as the same field that the sound source would cause in a free field.

2.1.8. Far field

The other part of the sound field that has been affected by the room properties is referred to as the far field (Bodén et al, 2012)). In the far field, every time the distance from the sound source is doubled the sound pressure decreases with 6 dB, given that the sound source is in a free field (which is the case in an anechoic chamber) (Nilsson et al, 2008). The total sound field of a room can be obtained by adding the energy density of the two sound fields, this is done with equation below (Bodén et al, 2012)).

$$L_p = L_W + 10 \cdot \log \left(\frac{Q}{4\pi r^2} + \frac{4}{A'}\right) \tag{2.6}$$

where
L_p is the sound pressure level
L_w is the sound power level
Q is the directivity factor of the sound source

r is the distance from the sound source
A' is the equivalent absorption area, it can be calculated with Sabine's formula

The distance from the source where the energy in the near field and the far field is equal is known as the critical distance. An expression for the critical distance can be obtained by making the two terms in the parenthesis in equation 2.6 equal.

$$\frac{Q}{4 \cdot \pi \cdot r^2} = \frac{4}{A'} \rightarrow r = \sqrt{\frac{A' \cdot Q}{16 \cdot \pi}} \tag{2.7}$$

2.1.9. Speech clarity (C_{50})

A listener notices not only the direct sound but also several delayed reflections (Vorländer, 2008). The sound that reaches the listener first without any obstacles is called direct sound and determines the perceived direction of the sound. The direct sound is followed by early reflections, these are the reflections that reaches the listener within 50 ms (ISO, 2009). Early reflections contribute to the direct sound impression by enhancing loudness, supporting speech intelligibility and the clarity of music (Vorländer, 2008). Reflections are delayed due to a longer path of sound propagation compared to direct sound. Reflections delayed more than 50 ms are referred to as late reflections and create the impression of reverberation, they can be perceived as disturbing (Vorländer, 2008).

Speech clarity compares the sound energy in the early reflections with the late reflections, it is expressed in dB (ISO, 2009). It can be calculated according to equation 2.8 below.

$$C_{50} = 10log \frac{\int_0^t p^2(t)dt}{\int_t^\infty p^2(t)dt} \tag{2.8}$$

where
t is the early time limit, in the case of speech it is usually 50 ms
p(t) is the instantaneous sound pressure of the measured impulse response

Speech clarity describes the quality of the speech transmission from the speaker to the listener. If a room has a lot of hard surfaces that causes increased reverberation, this can make it more difficult to perceive speech.

Figure 2.2. Direct sound (Black arrow) and early reflections (Grey arrows) (Ecophon, n.d.).

2.1.10. Speech intelligibility

Speech intelligibility is defined as the ability to perceive speech, it is an important measure of the effectiveness or adequacy of a communication system or to communicate in a noisy environment (Brüel & Kjaer, 2013). Speech intelligibility is usually expressed as a percentage of words or sentences, spoken by a speaker, that is correctly identified by a listener. To perform this kind of intelligibility test is however very time consuming and expensive, it is possible to determine the speech intelligibility through measurements instead. Measurements are performed with a small loudspeaker or a similar sound source and a microphone placed at the listener's position.

There are several parameters that effect the speech intelligibility, because of these the signal that is heard by the listener is never exactly the same as the one transmitted by the speaker. These parameters are:

- The level of background noise
- The distance between the listener and the speaker
- The loudness of the speech
- The voice spectrum of the speech
- The amount of reverberation present.

The most commonly used parameters for measuring the speech intelligibility are RASTI, STI and STIPA. RASTI and STIPA are simplified versions of STI (Brüel & Kjaer, 2013). RASTI (Room Acoustic Speech Transmission Index) describes how the sound is perceived by the ear. It is a measurement of how well speech is perceived in a room (Åkerlöf, 2001).

An analysis of speech shows that vowels are between 300 Hz and 3 kHz, voiced consonants between 300 Hz and 4 kHz and unvoiced consonants between 2,5 kHz and 12 kHz (Nilsson et al, 2008). The consonants contain most of the linguistic information, a hearing damage in the high frequencies will therefore lead to a larger impairment of the speech intelligibility compared to a hearing damage in the low frequencies.

2.1.11. Speech transmission index (STI)

STI is the most important and comprehensive parameter for measuring the speech intelligibility (Brüel & Kjaer, 2013). It is measured over seven octave bands from 125 Hz to 8 kHz. Table 2.1 shows the relationship between the speech transmission index and the speech intelligibility.

Table 2.1. The relation between the STI and the speech intelligibility (Brüel & Kjaer, 2013).

STI	0.00-0.30	0.30-0.45	0.45-0.60	0.60–0.75	0.75-1.00
Speech intelligibility	Bad	Poor	Fair	Good	Excellent

When the speech transmission is perfect the STI equals 1, for a normal classroom the value of STI should be greater than 0,75 (Ecophon, n.d.). Background noise, a long reverberation time and echoes are factors that contribute to a lower STI index.

2.1.12. Sound strength (G)

Because of reverberation, the room contributes to the sound or noise level from a sound source (Ecophon, n.d.). The sound strength is described as the ratio between the sound energy measured at a given position in a room and the sound energy measured at a distance of 10 meters from the sound source with free field conditions (which can be achieved in an anechoic chamber), with the sound coming from the same source (Beranek, 2011). By measuring the sound level in a real room and in an anechoic room with the same sound source, the room's contribution to the sound level can be determined (Ecophon, n.d.), see Figure 2.3.

The sound strength is strongly dependent on the absorption of the room, the more absorption in the room, the lower the value of G will be. In a free field, where there is total absorption of the sound, the room will not contribute to the sound level and G will be equal to zero.

a) b)

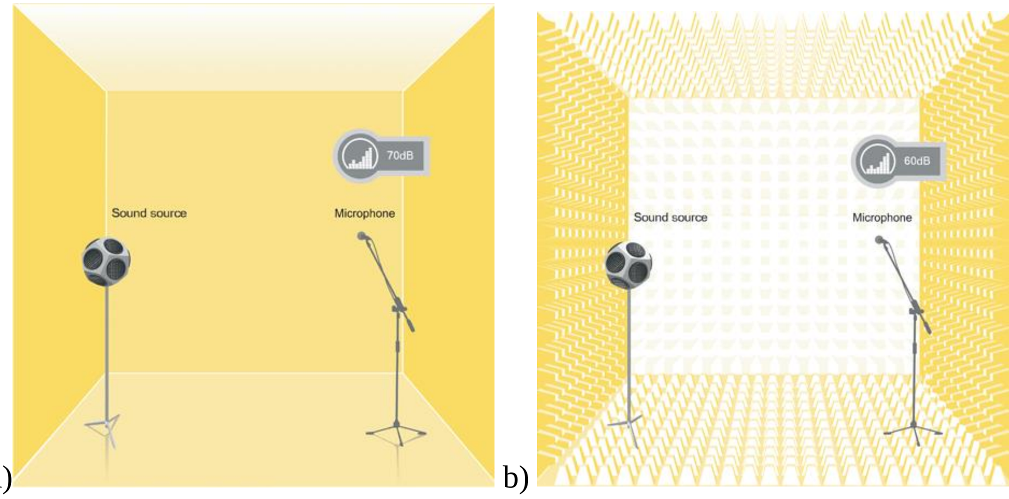


Figure 2.3. Measurement of sound level in a real rom (a) and an anechoic room (b) (Ecophon, n.d.).

2.2. The human hearing system

To get an understanding of how binaural recordings work it is important to first have a knowledge about how the human hearing system works. Here the function of the ears and the human hearing will be presented along with an introduction to psychoacoustics and some subjective parameters.

2.2.1. The ears

The ear consists of three parts, the outer, middle and inner ears, sound waves travel from the outer ear, through the middle ear and into the inner ear (Xie, 2013). The outer ear consists of the pinna and the ear canal, its most important function is to capture sound and lead it further into the ear (Vårdguiden, 2005). The pinna and the ear canal do not only collect sound but amplifies it as well (Resonance hearing clinic, 2013). The ear canal amplifies sound in the high frequencies, 2–4 kHz typically for an adult, how large the amplification becomes is individual and depends on for instance the length, volume and curvature of the ear canal. Usually the amplification in the higher frequencies becomes higher the smaller the ear canal is. Small children with small ear canals are therefore typically more disturbed by loud high frequency sounds. The ear canal is approximately 25-35 mm long from its entrance to the eardrum (Howard & Angus, 2009).

Sounds that arrive at the entrance of the ear canal can either be direct sound or sound reflected by the pinna. The difference in arrival time between them is direction dependent, this is because the incoming sounds from different directions are likely to be reflected by different parts of the pinna (Kinsler, 2000).The pinna modifies the spectra of incident sounds in a way that depends on the angle of incidence in relation to the head, this is important for vertical localization of sounds and front-back ambiguity (Moore, 2004).

The eardrum forms the barrier between the outer and the middle ears, it is a light, thin and elastic structure that converts acoustic pressure variations to mechanical vibrations in the middle ear (Howard & Angus, 2009). The middle ear is an air-filled space, in it are the ossicles, which are three small bones (Vårdguiden, 2005). These bones are called the malleus, the stapes and the incus. The function of the ossicles is to transfer the mechanical vibrations of the eardrum to the oval window (Howard & Angus, 2009), which forms the barrier between the middle and the inner ears. The inner ear mainly consists of the cochlea, which is a fluid filled bony spiral with a shelled structure (Xie, 2013). Its function is to convert the mechanical vibrations into nerve firings that eventually is to be processed by the brain (Howard & Angus, 2009).

2.3. Psychoacoustics

Psychoacoustics deals with how the ear receives and interprets sound (Kinsler, 2000). The final consumer of room acoustics is usually the listener that listens to for instance a concert or a speaker (Kuttruff, 1991). The listener does not have any requirements of the reverberation time or any other measurable quantities, instead he expects the acoustics of the room to support its purpose. The acoustics of a room should support music being played in it if it is intended for music or increase the speech intelligibility if it is intended for speech. The success or failure of an acoustic design of a room is to be decided by the judgement of all the users of the room in question. In the end it is the opinion of the listeners that decides whether the acoustics of a room is favorable or not. Even though measurements show that a room has good acoustics the listener is always right. If an audience can’t understand what a speaker is saying or perceives the sound of an orchestra to be too dry or too weak, the room has acoustic deficiencies. Since the listeners play a major role in determining the quality of the acoustics of a room some subjective quantities will be described in this chapter.

2.3.1. Loudness

The human ears can perceive sounds levels between 0 dB and 140 dB and frequencies between 20 Hz and 20 kHz, the sensitivity of the ears vary with the frequency of the sound (Bernström, 1987). The ears are less sensitive to lower frequencies, as a result of this a higher sound level is required for low frequencies to be perceived as loud as high frequencies (Bernström, 1987).

According to Moore (Moore, 2004) loudness is defined as the attribute of auditory sensation in terms of which sounds can be ordered on a scale from quiet to loud. Loudness is a subjective quantity and can therefore not be measured directly, instead listeners have been used in different experiments to evaluate loudness. The loudness level indicates how intense a 1000 Hz tone must be in order for it to sound equally loud. To determine the loudness level a listener is asked to change the intensity of a 1000 Hz tone until it is perceived to have the same loudness as the test sound. The level of the 1000 Hz tone that is perceived to have equal loudness as the test sound is the loudness level of the test sound, it is measured in the unit phon. If instead the 1000 Hz tone is fixed in level and the test sound is adjusted to achieve equal loudness, and this is repeated for several different frequencies, an equal-loudness contour is created.

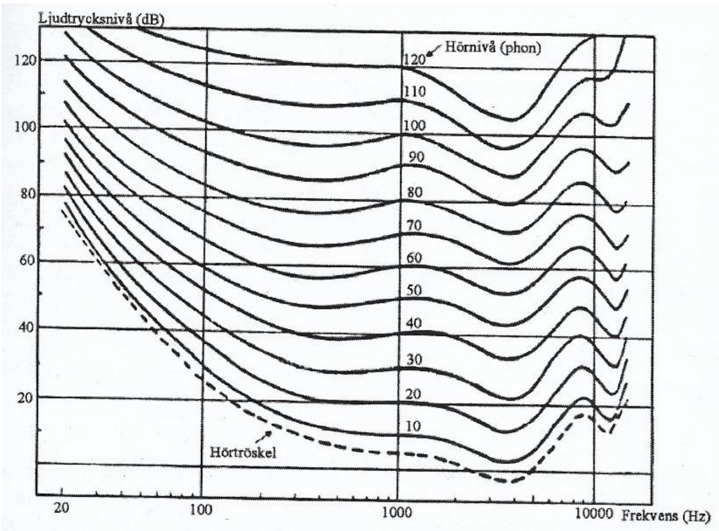


Figure 2.4. The contours of equal loudness. (Nilsson et al, 2008)

The value on each of the curves represents the phon-value which agrees with the sound pressure level at 1000 Hz (Kuttruff, 1991). The contours of equal loudness are measured for frontal incidence of the sound waves, it is then averaged over a number of subjects. The subjective loudness of a tone depends on the frequency of the tone and its sound level. Tones with different frequencies but with equal loudness often have different sound levels.

2.3.2. Pitch

According to Moore (Moore, 2004), pitch can be defined as the attribute of auditory sensation in terms of which sounds can be ordered on a musical scale. A change in pitch will result in the sensation of melody. Since pitch is a subjective quantity it is not possible to measure it directly, instead it is often investigated by asking listeners to adjust a tone so that it is perceived to have the same pitch as a complex sound. The pitch of pure tones does not only depend on the frequency of the sound but also on other acoustic parameters such as the sound pressure level (Zwicker & Fastl, 1999). The pitch of tones below 2 kHz decreases with increasing sound level while the pitch of tones above 4 kHz increases with increasing sound level (Moore, 2004). Since this is a subjective measurement there are considerable individual difference concerning the size of the pitch shifts with level and the direction of the shifts.

2.3.3. Timbre

According to Moore (Moore, 2004) timbre is defined as the attribute of auditory sensation in terms of which a listener can judge that two presented sounds, with the same loudness and pitch, aren't similar. It is the differences in timbre that makes it possible to distinguish between the same note played on different instruments such as piano or violin. Timbre relates to the quality of sound.

2.3.4. Spaciousness

Spaciousness is the acoustical feeling of space that a listener usually experiences in a room (Kuttruff, 1991). The sound in a closed space, as for instance a room, reaches a listener from different directions, the human hearing can't locate all the directions separately but processes them into an overall impression, this results in the sensation of spaciousness.

2.4. Anechoic chamber

For spaces that are intended for certain free field measurements, such as calibration of a microphone or for some psychoacoustic experiments, the accuracy and the reliability of the results would be impaired by the interference of the direct sound with the sound reflected by the room boundaries (Kuttruff, 1991). To avoid reflections a so-called anechoic chamber can be used. The usual requirement is that the walls of an anechoic chamber should have an absorption coefficient of at least 0,99 for all angles of incidence.

The most common construction of anechoic chambers is a rectangle shaped room with sound absorbers attached to all of the six inner surfaces (Schøyen Nielsen,). The floor is usually made of a net that consists of high-tension wires that is fastened in a solid frame. Since all the equipment needed for a measurement also has to be placed on the net floor there is a risk that the equipment will be shaken during measurements performed in the room. Since this is a fully standardized and specified measurement environment it gives everybody the possibility to perform the exact same measurements, since it is the same everywhere the results are also comparable.

2.5. Semi-anechoic chamber

In semi-anechoic chambers sound absorbers are usually attached to all surfaces except from the floor (Schøyen Nielsen). The floor in these types of rooms are often hard and reflective.

3. Binaural technology

Before discussing the technology behind binaural recordings, it is important to first have an understanding of how the binaural hearing works. In the beginning of this chapter the definition and characteristics of the binaural hearing will be discussed. This is followed by the fundamentals of the binaural recording technique. Finally, different methods for reproduction of binaural sound signals will be discussed along with the advantages and disadvantages of each method.

3.1. Binaural hearing

Binaural hearing is the ability to hear with two ears, this offers numerous advantages compared to hearing with only one ear, monaural hearing (Blauert, 2013). The differences in the input signal between the two ears provides information that makes it possible for humans to localize sounds. By processing the signals from both ears human hearing can recognize different directions and distinguish different sound sources (Head acoustics, n.d.). The ability to determine an actual or perceived spatial position of a sound source is referred to as auditory localization (Xie, 2013). The directions of incident sounds are usually determined in relation to the head of the listener or an artificial head. Because of this, three planes have been defined, the horizontal plane, the median plane and the frontal plane (Moore, 2004).

The intersection of all three planes is located at the center of the head, it defines the origin of a coordinate system that is used for determining the angles of incident sounds in relation to the head (Moore, 2004). The angles of the incident sounds can be defined by their azimuth and their elevation, sounds located on the median plane has an azimuth of 0 degrees, sounds located on the horizontal plane has an elevation of 0 degrees. Sounds with an azimuth and an elevation of 0 degrees is located straight in front of the head.

3.1.1. Cocktail party effect

The cocktail party effect describes the auditory system’s ability to focus on a desired speech sound source even if one or more interfering sound sources are present at the same time (Xie, 2013). The auditory system takes advantage of the spatial separation of the desired speech and the interfering speech sound sources to more effectively detect the wanted speech.

3.1.2. Precedence effect

For two coherent sounds the auditory system perceives a sound as if it is coming from the direction of the leading sound source when the difference in arrival time is within a certain limit, this is called the precedence effect (Xie, 2013). If the difference in arrival time between the sounds exceeds the limit a separate echo is perceived.

3.2. Binaural cues

Binaural hearing makes it possible for humans to localize sounds (Vorländer, 2008). There are mainly two mechanisms that affect the perception of the direction and distance of a sound source (Rumsey & McCormick, 2010). These include the detection of time or phase difference between the two ears, and of amplitude or spectral differences between the ears. If the sound source isn't located straight in front of the listener, the sound signals arrive at different times at the two ears and with different amplitudes. For instance, sound from a sound source located at the right side of the head has a longer distance to travel to the left ear than to the right, due to diffraction caused by the head and pinna the amplitudes of the sound signals at both ears display frequency-dependent differences (Kuttruff, 1991). A sound source that is not located straight in front of the listener will result in a difference in arrival time between the signals arriving at the ears of the listener. This is related to the angle of incidence of the sound (Rumsey & McCormick, 2010).

3.2.1. Interaural time difference (ITD)

Interaural time difference is the difference in time of arrival between the sound waves at the right and left ears (Xie, 2013). ITD is important for directional localization. When the distance between the sound source and the ears is identical for both ears ITD is approximately equal to zero, which is the case in the median plane. If the sound source deviates from the median plane, the distance to the sound source is different for each ear and ITD no longer equals zero. ITD is related to the anatomical dimensions of the head and therefore varies between individuals, it could be estimated from a pair of head-related transfer functions (see chapter 3.3).

3.2.2. Interaural level difference (ILD)

Interaural level difference describes the attenuation of the sound pressure level at the ear furthest away from the sound source when the sound source deviates from the median plane (Xie, 2013). This occurs because of the shadowing effect of the head. For low frequencies, where the wavelength is large compared to the size of the head, there will be little or no shadowing effect at all caused by the head (Moore, 2004). This is due to diffraction, since the wavelength is large compared to the size of the head the sound waves will bend very well around the head. The opposite occurs in high frequencies, here there is almost no diffraction and the shadowing effect of the head will be more apparent. ILD is both direction- and frequency dependent (Xie, 2013).

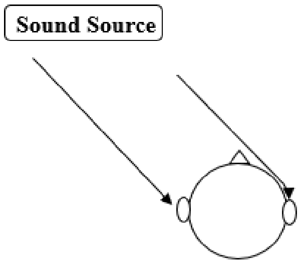


Figure 3.1. The path length from the sound source to the right and left ear.

3.2.3. Cone of confusion

The ITD and ILD are regarded as two dominant localization cues (Xie, 2013). However, a set of ITD and ILD is not sufficient for determining the unique position of a sound source. There are an infinite number of spatial positions with identical path lengths between the sound source and the two ears. If the shape of the human head is simplified and instead regarded as a pair of holes separated by a spherical obstacle, the holes will form the cone of confusion (Moore, 2004). Any sound source located the surface of this cone will give rise to the same ITD, in this case ITD alone is not sufficient to determine the unique position of a sound source. An infinite number of points exists in space within which ILD is the same for all points (Xie, 2013). For a human head, without simplified shape, the ITD and ILD is still not sufficient for determining the unique spatial position of a sound source since they do not vary monotonously with the position of the source. For example, in the median plane the sound pressures are almost identical for the two ears, ITD and ILD is thus equal to zero. For two sound sources located at mirroring positions in the front and back of the head, the resulting ITD and ILD will be the same for both positions of the sound source. This means that ILD and ITD only determines the cone of confusion in which the sound source is located, but not the unique spatial position of it.

In cases where two sound source positions give rise to identical ITD and ILD it is not possible to distinguish between the two sources in means of ITD and ILD (Xie, 2013). By rotation of the head around the vertical axis a change in ITD is achieved. Since head rotation could cause a change in, not only in ITD, but in ILD and the sound spectra at the ears as well, it too is an important cue for localization.

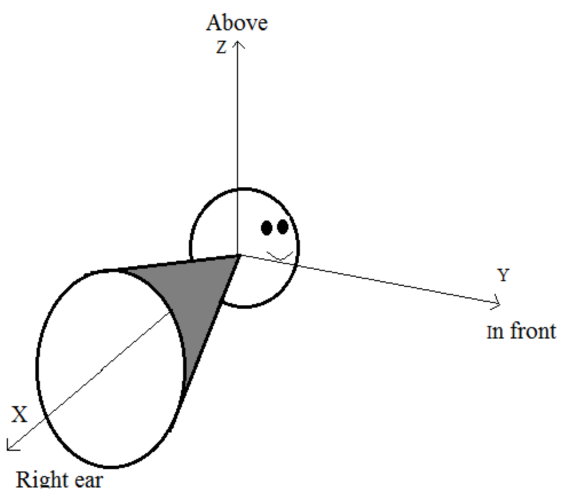


Figure 3.2. Cone of confusion, all sound sources located on the surface of the cone will give rise to identical ITDs.

3.2.4. Spectral cues

The human auditory system can determine the direction and distance of incoming sounds not only through the earlier mentioned interaural time difference and interaural level difference, but uses spectral cues as well (McKenzie et al, 2019). ILD and ITD helps determine the horizontal direction of the sound while spectral cues help with vertical localization. Spectral cues are caused by acoustic disturbances of the sound waves such as diffraction and reflections off and around the torso, head and pinna.

3.3. Head-Related transfer function (HRTF)

The head-related transfer function describes the effect that the listener's anatomical structures such as the head, torso and pinna, has on the sound (Xie, 2013). It varies between individuals since the anatomy of each human individual is unique. The distance between the ears and the shoulders has the biggest influence on the HRFT, it is also the parameter that varies the most with growth (Fels, 2008). This means that there are differences in the HRTFs not only between adult individuals, but also significant differences between adults and children.

For each location of a sound source a pair of HRTF exists, one for each ear (Xie, 2013). The most accurate and common way to obtain the HRTFs is through measurements, to measure the HRTFs of a human subject makes it possible to obtain individual HRTFs. The measurements are performed in an anechoic chamber where there are free field conditions. During the measurements it is important that the subject is absolutely still, if the human subject makes small movements of the head and body it can lead to measurement error. This is a larger risk during long-term measurements. Human subject can also unconsciously generate noises, this can damage the measurements. Because of the environment in the anechoic chamber and the fact that the subject must remain motionless during the measurement, this kind of measurement is not suitable to perform on children, instead an artificial head can be used (Fels, 2008). The measurement error of an artificial head is generally small. Some types of artificial heads are constructed with a simplified head and pinna, because of this the localization parameters associated with the details of the head and pinna may be lost in the resultant HRTFs. HRTFs of artificial heads are often measured at the equivalent position of the eardrums (Xie, 2013).

3.4. Binaural recording technique

With binaural technique it is possible to record and play back a sound in such a way as if the listener was present in the original sound field (Head acoustics, n.d.).The idea behind binaural technique is that the input to the hearing consists of two signals, these signals are the sound pressures at each of the two eardrums (Hammershøi & Møller, 1992). If these signals are recorded in the ears of a listener and reproduced exactly as they were, it is assumed that the complete auditive experience could be reproduced, including timbre and spatial cues.

The most natural way to obtain binaural recordings is to place a pair of microphones either at the entrances of the ear canals, or at an arbitrary point along the ear canals, of a human subject

(Xie, 2013). This way the influence of the human anatomy on the sound is preserved. However, the slightest movement or noise from a human subject will affect the accuracy of the recordings, and thus how well they reflect the original sound field. To avoid the influence on the sound that can be caused by a human subject an artificial head (dummy head) can be used to do binaural recordings. In this case microphones are placed in the same places as described above for human subjects. An artificial head simulates the anatomical structures of real humans and is made of specific materials with acoustic properties similar to those of humans. Since binaural signals are influenced by the anatomical structures of humans they vary between individuals. Therefore, the performance of the reproduction of binaural signals recorded with an artificial head depends on how similar the anatomical structures are between the artificial head and the listener. The more similar they are the more authentic the binaural recordings will be perceived. Because of this the perception of the binaural recordings will vary between different listeners (Xie, 2013).

3.5. Impulse responses

An impulse response is the response of the room on excitation with an impulse (Vorländer, 2008). It gives a description of the changes a sound signal goes through as it travels from one point in the room to another (Kuttruff, 1991). Many room acoustic parameters, such as reverberation time, speech clarity, etcetera, can be derived from the impulse response. Impulse responses can either be measured or synthesized using auralization techniques (Cohen Tenoudji, 2016).

3.5.1. Room impulse response (RIR)

The sound transmission between a sound source and a receiver can be described by the room impulse response (Xie, 2013). A lot of room acoustic parameters can be evaluated through RIR since it contains temporal characteristics of both direct and reflected sounds. In reality when there is a human listener the situation becomes more complicated. This is because in this case there are two receivers instead of one, the ears, and the human anatomy modifies the sound reaching the two ears through reflection and diffraction.

3.5.2. Head-Related impulse response (HRIR)

Head related impulse response is the impulse response from a point sound source to both ears in a free field (Xie, 2013).

3.5.3. Binaural impulse response (BRIR)

Binaural impulse responses are binaural pressure signals in a room, which contains spatial information of both sound source and environment (Xie, 2013). BRIR depend on the positions of the sound source and the receiver.

3.6. Binaural sound reproduction

Binaural signals can, as previously explained, be obtained by binaural recordings (Xie, 2013). If the sound pressures at the eardrums generated when binaural signals are reproduced via headphones are exactly equal to the sound pressures that occurred in the original sound field, the sound event can be reproduced to the listener. Binaural signals can be recorded at any point between the entrance of the ear canal and the eardrum. To avoid getting the wrong sound pressures at the eardrums the measurement position must be taken into account before reproducing a binaural signal. This is because there is a difference between binaural signals recorded at different positions. Since results measured at different points differ from each other, and the length of the ear canal is individual, it is common that a point close to the ear canal entrance or the eardrum is selected, for the purpose of standardization.

Since the binaural signals recorded at different points along the ear canal differ from each other, it is not correct to directly render the recorded binaural signals without taking the measurement position into account. This can lead to incorrect pressures at the eardrums, the reproduced sound will in this case therefore not represent the original sound field.

Binaural recordings can be reproduced either by headphones or by loudspeakers.

3.6.1. Binaural sound reproduction through headphones

Binaural signals can be obtained by recording at a specific point along the ear canal of a listener. The sound pressure at the listener's eardrum can be generated by reproducing the binaural signals with the use of headphones. This sound pressure should be exactly the same as the sound pressure at the eardrums caused by the real sound source.

The headphone-to-ear-canal transfer function, HpTF, is defined as the influence of the headphones, on the binaural signals, introduced in the rendering process (Zhong, 2013). To ensure an authentic reproduction of binaural signals a headphone equalization must be performed (Xie, 2013). The headphone equalization is a process where the influence of the headphones is eliminated. If the binaural signal is equalized by the inverse of the HpTFs before being played back by a headphone the HpTFs could be eliminated, or at least reduced as much as possible.

HpTF contains both the headphone response and the acoustical coupling between the headphone and the external ear. Because of the individual differences in the shapes and dimensions of the external ear the HpTF varies between individuals. The HpTF refers to the transfer characteristics from the electric input signal of the headphone to the pressure at a reference point somewhere along the ear canal of the listener (Xie, 2013). If the transfer characteristics is different between the right and the left ear (or microphones) the headphone equalization must be done separately for both ears.

The headphone-to-ear-canal transfer function (HpTF) can be calculated as:

$$Hp(f) = \frac{P'_x}{E} \tag{3.1}$$

where:
$Hp(f)$ is the headphone-to-ear-canal transfer function
P'_x is the reference point along the ear canal
E is the binaural signal

If the eardrum is chosen as the reference point, $Hp(f)$ is instead referred to as the headphone-to-eardrum transfer function (HETF) (Xie, 2013).

The HpTFs of circumaural headphones are strongly dependent on the size and shape of the external ear, because of this they are different between individuals just like the HRTFs (Xie, 2013). Because of this individual HpTFs should be used for headphone equalization to ensure an accurate replication of the binaural signals at the eardrum.

A playback via headphones alone may be insufficient since some headphones can't generate high enough sound pressure levels in the low frequency range (Head acoustics, n.d.).

Directional error

In headphone reproduction, an accurate virtual sound source can be perceived by the listener if the sound pressures created by a real sound source at each of the eardrums of the listener is replicated precisely (Xie, 2013). However, sometimes subject-dependent directional errors occur, examples of these errors include:

- Reversal error (front-back confusion). The position of a virtual sound source in the front hemisphere is perceived to be in the mirroring position in the rear hemisphere, or the reverse. Confusion regarding if the position of the sound source is up or down sometimes also occurs, this is referred to as up-down confusion.
- Elevation error. The elevation of a virtual sound source in the front median plane is perceived to be elevated to a higher position.

As mentioned earlier, ILD and ITD only determines the cone of confusion in which the sound source is located, but not the unique spatial position of it. The spectral cues caused by the pinna and the localization cues caused by head movement are important for clarification of the front-back ambiguity and vertical localization. However, binaural recordings obtained by using an artificial head usually lack the localization cues caused by the movement of the head since they are often performed with the head standing still. The spectral cues caused by the pinna is individual dependent since the anatomy of the pinna varies between individuals. These cues are therefore often sensitive for impairment that could occur if an artificial head with dimensions that does not match the anatomy of the listener is used. These types of cues will also be damaged if individual HpTFs is not used during headphone equalization or in cases where no headphone equalization is performed at all.

In-head localization

When binaural signals obtained by binaural recording technique are reproduced through headphones, an auditory event often occurs inside of the head leading to an unnatural hearing experience (Xie, 2013). The virtual sound source is often perceived to be located on the surface of the head rather than inside of it, this is often the case if the actual sound source is located directly in front of the artificial head during recording. This is referred to as in-head localization, the elimination of it is called externalization. In-head localization is believed to be caused by incorrect spatial information in the sound reproduction process. The human hearing system is likely to create the illusion that the virtual sound source is located inside the head when the binaural signals fails to provide localization cues identical to those of the real source.

3.6.2. Binaural sound reproduction through loudspeakers

Since binaural signals were originally intended for headphone reproduction additional signal processing is needed in order to produce binaural signals through loudspeakers (Xie, 2013). To directly render binaural signals through loudspeakers causes timbre coloration, this could be compensated for by free-field equalization. In rooms the reflection field will dominate when the recording device is located far from the sound source. In this case the timbre coloration that occurs in loudspeaker reproduction can be taken into account by diffuse-field equalization.

When binaural signals are reproduced through two loudspeakers located in front of the listener, one for each ear, crosstalk between each loudspeaker and the ear opposite of it occurs, this causes an impairment of the directional information in the binaural signals. Crosstalk cancellation means that the binaural sound signals are pre-corrected or filtered to cancel sound transmission between each loudspeaker and its opposite ear before the binaural signals are reproduced through loudspeakers. Binaural signals reproduced by loudspeakers with crosstalk cancellation taken into account can generate the same binaural pressures as those that are produced by a real sound source. Crosstalk cancellation shows a position dependent performance. For a given loudspeaker configuration crosstalk cancellation is only effective within a limited region, if the position of the head deviates from the optimal position the crosstalk cancellation will be affected negatively, and the binaural pressures will be altered.

4. Listening test

Listening tests are a good way to gain information about the subjective experience of a sound. It is important to remember that the final consumer of room acoustics is always the listeners that are going to use the room in question. There are several methods for performing listening tests, in this thesis scales were used. To allow the listeners to score the different sounds in the listening tests has the advantage that the scores can be treated roughly as measurements (David, 1963). If the grades are hard to define or if the differences between the presented sounds are small the method of rating could be less successful. An advantage with using scales is their ability to produce absolute values (Johansson, 2005). However, scales can be hard for the listeners to use, furthermore uncertainties whether the listeners have understood and used the scales equally can also occur. The scoring could also be affected by the memories and experiences of the listeners, this could result in different listeners making different ratings (Söderholm, 1998).

4.1.1. Before the listening test

Before the listening test is performed it is important that the participating listeners are instructed properly (Head acoustics, n.d.). The instructions should include all necessary information about the planned listening test, it could also be a good idea to explain the purpose of the listening test. However, it is important that this information is only given if it is not likely to affect the listeners judgements. The instructions should be given in both written and oral form, it is often sufficient to use a summary of the oral instructions for the written ones. It is also important that the oral instructions are the same for all participating listeners. The listeners could also be informed about the duration of the test so that they know what to expect. It is only if the listeners have fully understood the task that they will feel confident during the test and complete it with reliable results.

After the instructions have been given a training of the listeners could be performed. If such training is needed depends on the difficulty of the task and the experience of the listeners. Such training consists of the listeners listening to some, or all, of the sounds in the listening test in advance. This would lead to the listeners being more prepared for the listening test and they would know what to expect. It is important that the training is not too long since this could cause a loss of concentration or motivation among the listeners.

4.1.2. During the listening test

After proper instructions and training have been given, the listening test can begin. It is important that the listeners are not disturbed during the test, they should also not feel left alone (Head acoustics, n.d.). The test conductor should be available in person or by phone. During the listening test, the judgements of the listeners can be written down on prepared forms, but often the judgements are entered directly into the computer by using suitable software. By using suitable software errors that occurs when the data is transferred from paper into the computer can be avoided.

A listening test should not last longer than 45 minutes in order to ensure that the listeners do not lose concentration. The number of sounds and their length should be chosen so that this time limit is not exceeded. After the test, the test conductor should ask the listeners about their impression regarding the test procedure, the duration, etc.

4.1.3. Test environment

The test room should have sufficient ventilation and a pleasant temperature to ensure that the listeners feel comfortable (Head acoustics, n.d.). It is important that the background noise is low, if sounds with a low volume is to be judged the test has to be conducted in a soundproof room.

The closer the conditions of the test environment are to the conditions the sounds are recorded in the more meaningful the results of the listening test will be. If the characteristics differ significantly between the recording room and the test room the inconsistency between the visual and acoustic impression can lead to untrained listeners perceiving the sounds as, for instance, too loud.

4.1.4. Sounds

It is important that the sounds used in a listening test have a high and consistent quality, it is recommended that an artificial head is used for recording to provide the listeners with spatial information (head acoustics, n.d.). It is recommended that the sounds are edited to ensure that they don't contain any unnecessary background noise and so that all sounds have the same length. If the sounds are different in some aspects it is difficult to know which aspects that has influenced the judgements of the listeners. Since some listeners can easily be distracted if the presented sounds have different loudness levels it is in some cases recommended that the sounds are edited so that they are all perceived as equally loud.

It must be decided if the sounds are to be played back through loudspeakers or headphones. If loudspeakers are used the acoustic properties of the test room should be taken into account to ensure that the listeners hear what they are supposed to hear. Playback through headphones is a good way to ensure that each listener hears the same signal.

The stimuli time has to be long enough for the participants to make ratings, but if it is too long it can have a negative effect on the listeners attention (Söderholm, 1998).

4.1.5. Creating a listening jury

The number of listeners and the selection of them has an effect on the results of the listening test (Head acoustics, n.d.). Therefore, they should be selected with care. A listener that has participated in listening tests is considered to be a trained listener, which means that the listener has knowledge about the procedure and the purpose of the listening test (Söderholm, 1998). An experienced listener will have no problem solving more difficult tasks that untrained listeners are not capable of (Head acoustics, n.d.). However, trained listeners can sometimes overrate certain aspects so that a sound is rated as unacceptable, while an untrained listener could have considered the same sound to be acceptable.

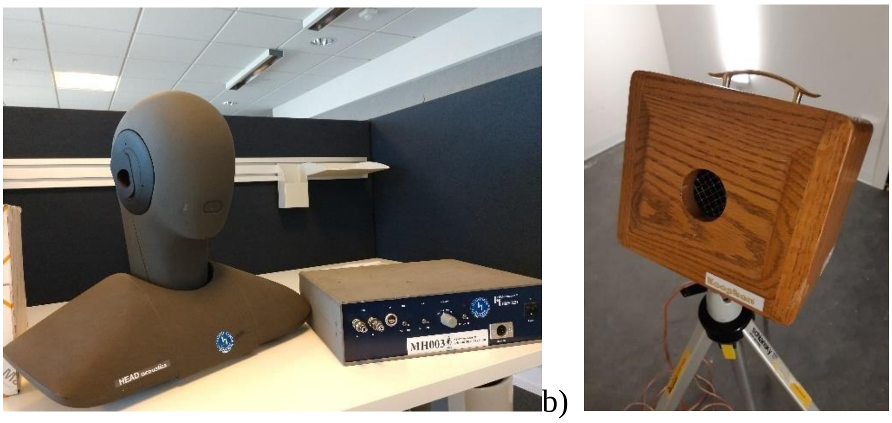
Apart from the previous experience of listening tests among the listeners, their memories and previous experiences should also be checked since this could affect their opinions of the presented sounds.

To perform listening tests in a group saves time since several listeners can participate at the same time, it also ensures that the conditions are identical for all the listeners (Head acoustics, n.d.). Listening tests in groups should however only be performed if the listeners can't disturb or influence each other. Sick listeners should be excluded from a group test since their coughing could distract the rest of the group.

5. Method

5.1. Binaural recordings

To perform the binaural recordings the head measurement system HMM II from Head acoustics was used. The HMM II consists of the HSU II artificial head along with an electronics unit, (..) and associated cables. As a sound source an open-office loudspeaker from Acculab was used, see Figure 5.1.



a) b)

Figure 5.1. The components included in the HMM II head measurement system, a) shows the artificial head along with the electronics unit, b) shows the loudspeaker used

5.1.1. The electronics unit

The settings of the electronics unit were set as following:

- “Ref” was set to off so that the reference tone at the output was switched off.
- “HP” was set to off to deactivate the high-pass filter.
- “LIN” was chosen as the equalization mode, this means that the signal from the artificial head was not equalized.
- “Level” was set to 114 dB in order to have high enough amplification without risking the peaks of the soundwaves to be cut off.

The electronics unit was connected to the HSU II through a 20 m long cable, and to a laptop through an AUREON XFIRE 8,0 HD soundcard.

5.1.2. Recording method

The HSU II was placed on a stand so that its mouth was approximately 1,20 m above the floor. It was placed at a distance of 2,0 m from the sound source, in this case a directional loudspeaker with dimensions similar to a human head. The loudspeaker was also placed on a stand so that its acoustical center had the same distance to the floor as the HSU II. The loudspeaker and the artificial head were placed so that the sound would have frontal incidence. Anechoic recordings of both male and female speech were then played through the loudspeaker. The binaural recordings were then performed by using the program MATLAB on the laptop. A total of two recordings (one for female speech and one for male speech) was performed in each room configuration. This recording procedure was repeated in exactly the same way in all of the different room configurations. The recordings were then saved as WAV-files to make the later editing of them easier. Due to problems with the saving of the recordings the recording of the male speech is missing for one of the room configurations.

5.1.3. Impulse responses

Impulse responses were measured for all room configurations except the anechoic chamber (Room E). A microphone was placed in the same position as the head had been, and a loudspeaker in the same position as the loudspeaker had been during the binaural recordings. The impulse responses were measured by using a sweep signal. This was repeated for all room configurations.

5.1.4. Room configurations

To give an answer to the problem of this thesis binaural recordings were performed in a number of different room configurations with a different amount of absorbents present in them. A total of six different room configurations were used, to create these the acoustic lab at Ecophon in Hyllinge was used. The room configurations were of different room volumes, they are all described below. Floorplan sketches over for each room configuration can be found in Appendix 1. In them the placement of the loudspeaker, the artificial head and any extra insulation are shown.

Room A

Room A could be assumed to represent an average classroom. The dimensions of the room were 7,6x7,3x2,9 m, which resulted in a volume of 160,9 m^3. In here absorbers of the type “Wall panel A” from Ecophon with dimensions 2,7x1,2x0,4 m was placed on two of the walls. The absorbers in the ceiling consisted of acoustic plates with a thickness of 15 mm, see Figure 5.1. The sound pressure level was measured with a sound level meter at a distance of 1 m from the loudspeaker, this showed a sound pressure level of 62,0 dBA for male speech and 62,2 dBA for female speech. In this room configuration two recordings were performed, on with female speech and one with male, they will be referred to as A1. After this both the loudspeaker and the artificial head was moved a short distance inside the room before another set of recordings were performed, this was done to investigate if the placement of the HSU II and the loudspeaker was critical to the resulting binaural recordings. Recordings in this room configuration will be referred to as A2.

Figure 5.2. The binaural recordings being performed in room A. Absorber of the type “Wall panel A” located at the ´back of the room (red wall panel).

Room B

Room B could be assumed to represent a small meeting room. This rooms had the dimensions 4,0x3,1x2,75 m, this resulted in a volume of 34,1 m^3. In this room two absorbers with a thickness off 100 mm and dimensions of 1,2x1,2 m and 0,6x1,2 m respectively was placed on two of the walls. The absorbers in the ceiling consisted of acoustic plates of the type “Advantage” from Ecophon and had a thickness of 15 mm. Recordings performed in this room configuration will be referred to as B1. When the recordings had been done in this room configuration, absorbers of the type “Wall panel A”, with dimensions 2,7x1,2x0,4 m, from Ecophon was placed in the room. One was placed on the wall behind the loudspeaker and one on the floor between the loudspeaker and the HSU II. After this another set of recordings were performed, they will be referred to as B2. Figure 5.2 shows room B before and after the extra absorbers were placed in the room.

a)  b) 

Figure 5.3. Room B without (a), and with (b) extra absorbers in the room.

Room C

Room C represents a semi-anechoic chamber, the dimensions of the room were 4,7x2,9x2,75 m which resulted in a volume of 37,5 m^3. In here all of the walls, even the door, were covered by a 100 mm thick layer of absorbers of the type "Industry modus" from Ecophon. The absorbers in the ceiling had a thickness of 40 mm and were of the type "Master A". On the floor a partly absorbing carpet was placed, since it was relatively thin it only absorbed high frequencies, see Figure 5.3.

a)  b) 

Figure 5.4. The binaural recordings being performed in room C (a), and the thickness of the absorbers placed on the walls and the door (b).

Room D

Room D represents a reverberation chamber. The dimensions of the room were 4,0x3,6x4,1 m, the volume was 59,04 m^3. In this room there are no absorbents present and the floor and the walls are covered with highly reflective materials, see Figure 5.4.

Figure 5.5. The reverberation chamber.

Room E

Room E is an anechoic chamber located at LTH, see Figure 5.5. In here all of the inner surfaced are covered with thick absorbers, even the floor. Instead the floor consists of a net. Before the recordings were performed the sound pressure level from the loudspeaker was measured with a sound level meter to verify that it was the same level as it was in the lab at Ecophon.

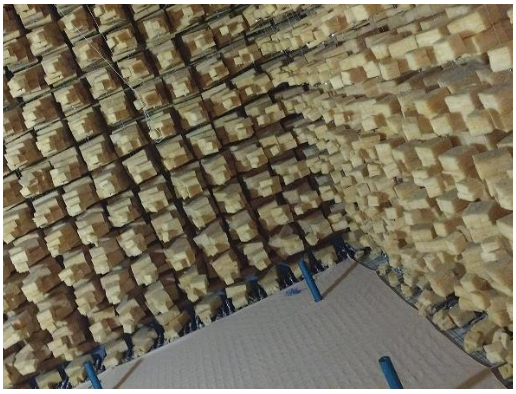
a)  b) 

Figure 5.6. The walls (a), and the floor (b), of the anechoic chamber.

5.2. Editing

After the recordings had been made the sound files were edited in the Audacity editing software (Audacity, 2018). Here all sound files were normalized so that they would have the same max peak level, the peak amplitude of all recordings was normalized to 0 dB. This was done to eliminate the difference in loudness between the recordings so that the listeners in the listening test would not focus on this. After normalization a noise could be heard in the background of all recordings. Because of this, noise reduction was performed on all sound files in order to try to eliminate the noise. This did however not work since it resulted in the noise not being perceived as equally loud in all recordings. Because of this no noise reduction was performed and the noise in the background was left in all recordings. Finally, all sound files were edited so that they would start at the same place and have the same length. The length of the sound files was chosen to 5 seconds.

By editing the recordings in this way, they would be similar in every way except for the amount of reflections present, which is the parameter of interest to investigate.

5.3. Listening test

A total of 15 people participated in the listening test, 9 of them are working in the field of acoustics. The listening test was performed individually to avoid the individual from getting influenced by the group. Since the test consisted of recordings performed in different types of rooms it was difficult finding a room to perform the test in that matched the recordings that the listeners were to listen to. The test was therefore conducted in a small meeting room, which was one of the types of rooms present in the test. At the moment for the test each listener was asked to fill in a small questionnaire about their hearing and their previous experience in listening tests, this questionnaire is attached in appendix 2.

In the listening test recordings of the same sounds (female or male speech) was compared to each other. Since the purpose of the thesis is to investigate how many absorbers that are needed before a room is perceived as an anechoic chamber there is no need to compare all the recordings to each other. All the recordings from all room configurations was therefore only compared to the anechoic recordings. This means that there was a paired comparison where one recording of each pair of sounds was an anechoic recording, see Figure 5.7.

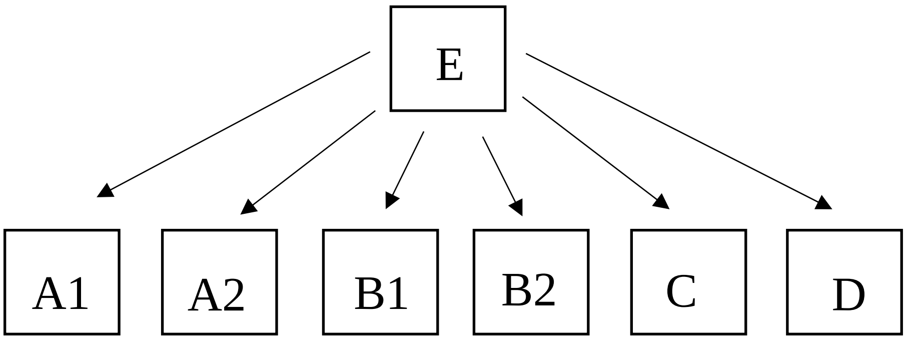


Figure 5.7. Sketch over how the recordings were compared.

The order of the pairs was randomized so that different listeners listened to the pairs of sounds in a different order. The binaural recordings were then played back through a laptop. The listeners then listened to the recordings through headphones, to make sure that the sounds were played back correctly a pair of headphones with a straight frequency response were used. The name of the headphones used was Sennheiser HD 600. After each pair of sounds presented the listeners answered a questionnaire about how similar they perceived each pair of sounds to be, they were asked to mark the similarity of each pair of sounds on a scale from 0-10. Before each listener performed the listening test, the test conductor explained the purpose and the procedure of the listening test, this information was also available to the listeners in written form as a part of the questionnaire. The listeners also got to listen to two of the recordings before the test started so that they would have an idea of how to relate to the scale. The questionnaire is attached in appendix 3.

5.3.1. Statistical evaluation

The scores from the listening test were compiled and a mean value and a standard deviation was calculated for each of the recordings. With the mean value and standard deviation known, a hypothesis test was performed using the method of t-test. What was of interest to investigate here was if the differences between the recordings that the listeners experienced in the listening test were significant. The mean value of each recording was therefore compared to the value 10, which was the highest possible score on the listening test, to see if there scoring suggests that there is a significant difference between the compared recordings. The hypothesizes was defined as below:

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

where
H_o is the null hypothesis
H_1 is the alternative hypothesis
μ is the mean value of each recording
μ_0 is the value that the mean value is being tested against

Then the t-value could be calculated for each recording by using the equation 5.1 below (Engstrand & Olsson, 2003).

$$t = \frac{\mu - \mu_0}{\sqrt{\frac{s^2}{n}}} \tag{5.1}$$

where
μ is the mean value
μ_0 is the value that the hypothesis is tested against, in this case 10
s is the standard deviation
n is the number of observations

5.4. The critical distance, *r*

As was shown in chapter 2.1.7. the critical distance can be calculated according to equation 2.7.

$$\frac{Q}{4 \cdot \pi \cdot r^2} = \frac{4}{A'} \rightarrow r = \sqrt{\frac{A' \cdot Q}{16 \cdot \pi}} \tag{2.7}$$

Where *A'* can be calculated with Sabine's formula.

$$A' = 0{,}16 \cdot \frac{V}{T} \tag{5.2}$$

A simple calculation was performed for all room configurations and all octave bands to check whether the artificial head was located in the far during the recordings. The directivity index of the loudspeaker, Q, was assumed to be equal to 1. Since a loudspeaker with a directivity similar to a human mouth was used, the directivity is higher than 1 in reality. Since these calculations were only performed as a quick check it is safe to assume the directivity index of the loudspeaker to be equal to 1. This will lead to an underestimation of the critical distance, which will be larger in reality.

5.5. The mean absorption coefficient, α_m

The mean absorption factor was calculated for each room with the rewritten version of Eyring's formula.

$$T = -0{,}16 \cdot \frac{V}{S_t \cdot \ln(1 - \alpha_m)} \rightarrow \alpha_m = 1 - e^{\frac{-0{,}16 \cdot V}{S_t \cdot T}} \tag{1.1}$$

where
V is the volume of the room
T is the reverberation time
S_t is the total surface of the room boundaries (walls, floor and ceiling)

The volume and the total surface of the room boundaries could be calculated with the room dimensions known. As mentioned earlier, the reverberation time was measured for each octave band in all of the room configurations. With this known, the mean absorption coefficient, α_m, was calculated for each of the octave bands. The single value of α_m was then calculated as the mean value of the values of the octave bands from 250 – 4000 Hz.

6. Results

6.1. Impulse responses

From the impulse response measured in each room configuration in the same place as the artificial head and the loudspeaker was placed, the sound pressure level, reverberation time and the speech clarity could be decided for every octave band.

6.1.1. Sound pressure level

As can be seen in Figure 6.1 the sound pressure level in room D is significantly higher than in the rest of the room configurations. For frequencies over 500 Hz room C has the lowest sound pressure level.

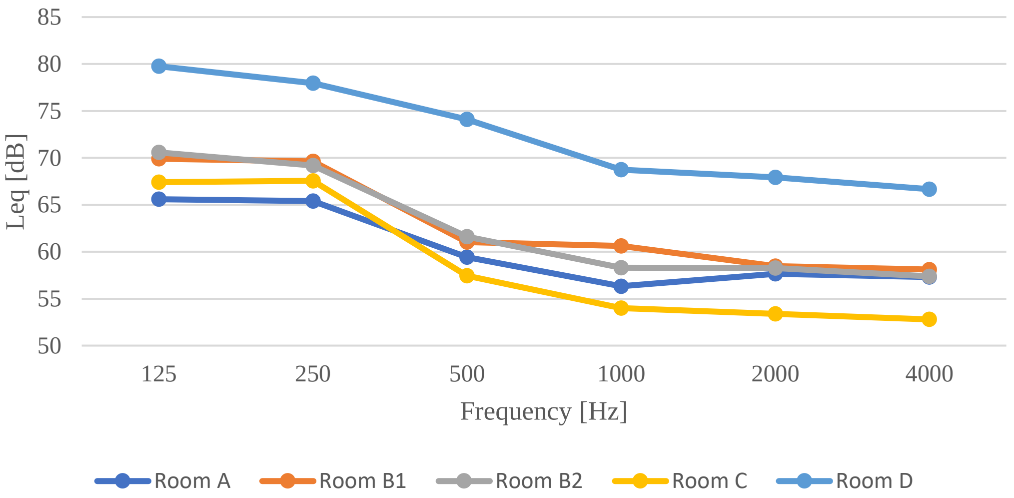


Figure 6.1. The sound pressure level in each room configuration for each octave band.

6.1.2. Reverberation time (T_{20})

As Figure 6.2 shows, the reverberation time in room C is significantly shorter in relation to the other room configurations, especially in the high frequencies. The reverberation time for room D was excluded in Figure 6.2 because it deviated significantly from the others, this was done to get a better resolution of the figure. As for room B, the graph shows that the reverberation time becomes significantly lower in the lower frequencies with the extra absorbers present. Above 2000 Hz room B1 and room B2 have a very similar reverberation time.

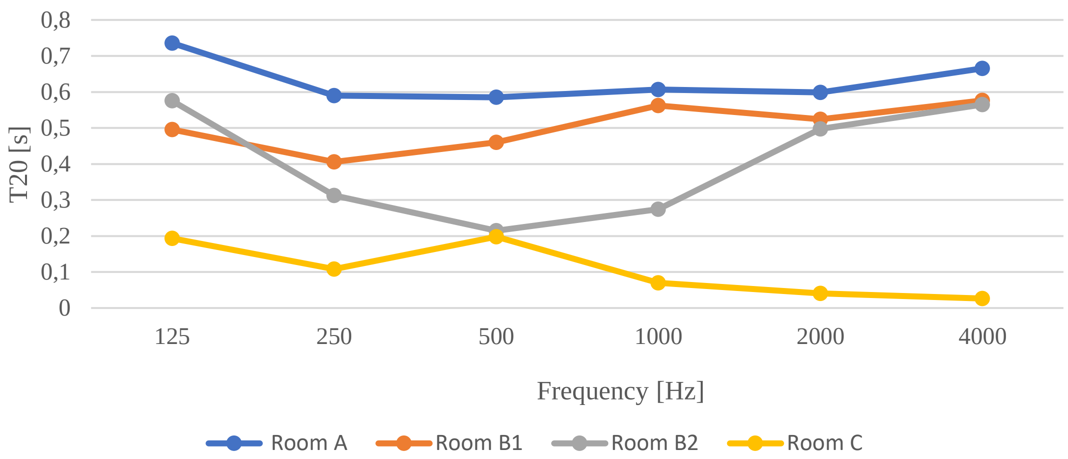


Figure 6.2. The reverberation time in each room configuration for each octave band.

6.1.3. Speech clarity (C_{50})

As shown in Figure 6.3 the highest value of C_{50} is achieved in room C, and for frequencies over 1000 Hz there is no value for it, this is due to a lack of late reflections. The lowest value of C_{50} occurs in room D, this is the only room configuration where the values are negative. The rest of the room configurations have very similar values of speech clarity.

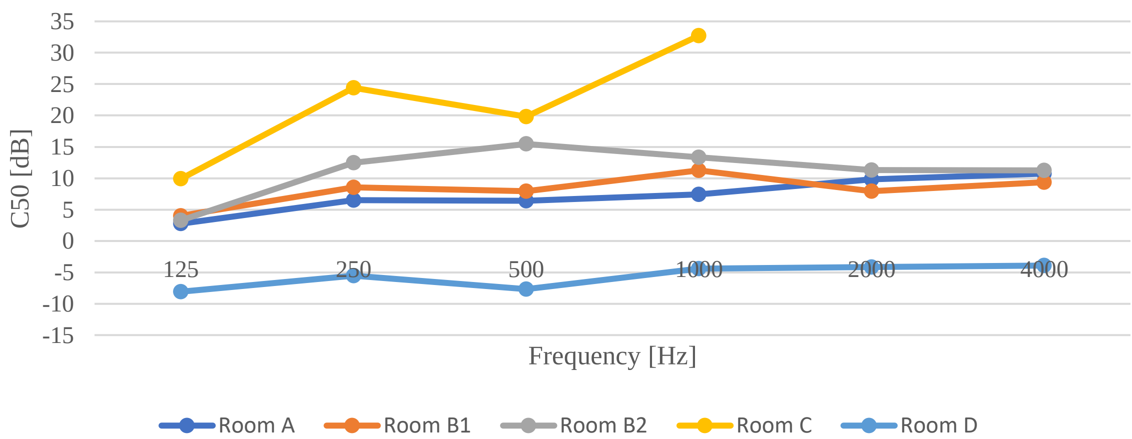


Figure 6.3. The speech clarity in each room configuration for each octave band.

6.2. Listening test

6.2.1. Questionnaire

All the 15 listeners in the listening test were asked to answer a small questionnaire about their hearing and their previous experience with listening tests, the results are shown in table 6.1 below.

Table 6.1. The listeners answers to the questionnaire about their hearing and their precious experiments of listening tests.

	Yes	No
Do you have any known problems with your hearing?	5	10
Do you consider yourself to have normal hearing?	11	4
Have you participated in listening tests before?	5	10

As table 6.1 shows, 5 of the listeners had a known problem with their hearing, 4 listeners didn't consider themselves to have normal hearing and 5 of them had participated in listening tests before. As earlier mentioned in chapter 5.3, 9 of the listeners are currently working in the field of acoustics.

6.2.2. Scores

In Figure 6.4 and 6.5 the scores of the 15 listeners on each of the recordings are shown for female and male speech respectively. This was done to try and identify any outliers, as the figures shows, no obvious outliers can be identified.

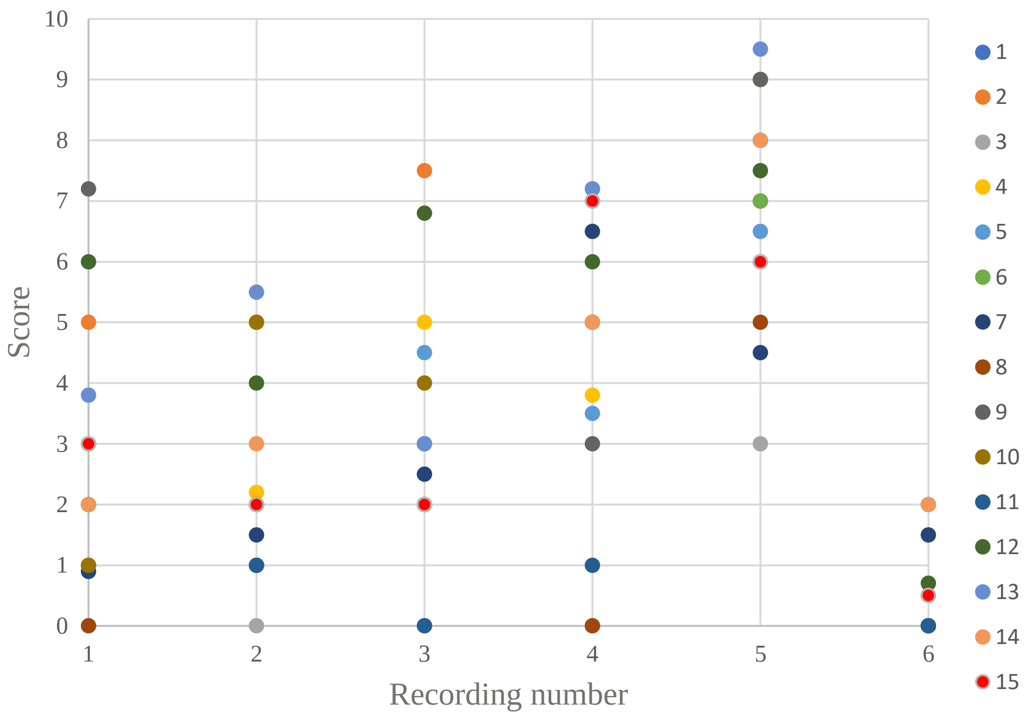


Figure 6.4. How the listeners scored the different recordings of female speech. Recording number 1=A1, 2=A2, 3=B1, 4=B2, 5=C, 6=D.

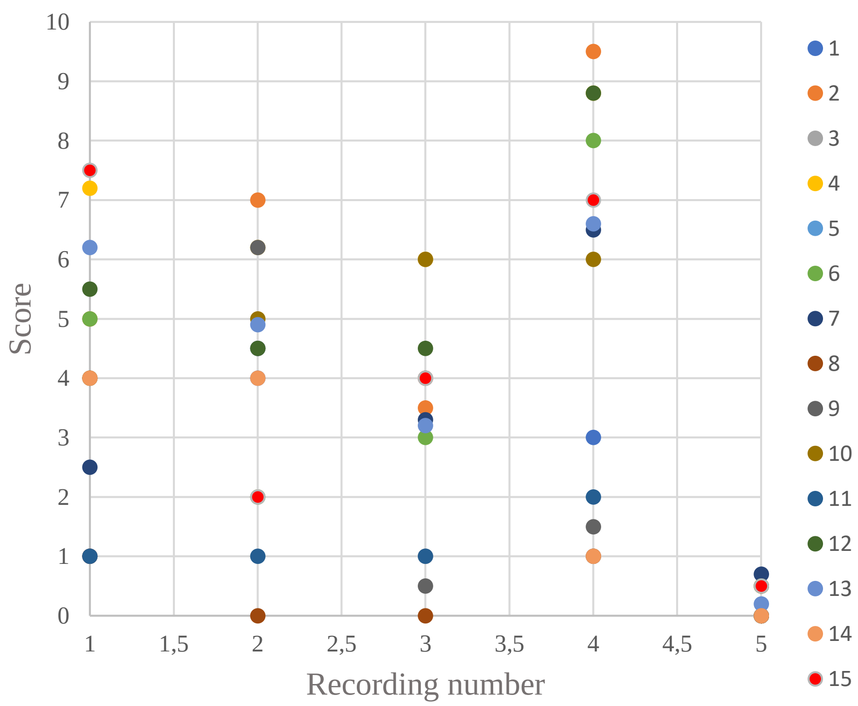


Figure 6.5. How the listeners scored the different recordings of male speech. Recording number 1=A1, 2=A2, 3=B1, 4=B2, 5=D.

6.2.3. Statistical analysis

After the listening test was performed the scores for each recording were compiled and the mean value and standard deviation for each recording were calculated, the results are shown in table 6.2 below.

Table 6.2. Mean value and standard deviation for each recording.

Recording	Female speech		Male speech	
	Mean value	Standard deviation	Mean value	Standard deviation
A1	2,79	2,04	4,06	2,10
A2	2,41	1,71	3,95	2,10
B1	3,09	2,16	3,10	1,83
B2	4,48	2,54	5,08	3,10
C	7,00	1,87	-	-
D	0,51	0,73	0,29	0,53

As table 6.2 shows, the mean value is slightly higher for the male speech in almost all cases. Room C has by far the highest mean value of all room configurations for the female speech.

With the mean value and standard deviation known, a hypothesis test could be performed using the method of t-test. The hypothesizes was as earlier demonstrated defined as:

$$\begin{cases} H_0: \mu = \mu_0 \\ H_1: \mu \neq \mu_0 \end{cases}$$

where
μ is the mean value of the scoring of each recording
μ_0 is 10 in this case

To test the hypothesis the method of t-test was used, the t-value for each recording was calculated according to equation 5.1. The results are demonstrated in table 6.3 below.

Table 6.3. The calculated t-value for all recordings.

Recording	Female speech	Male speech
A1	-13,71	-10,95
A2	-17,17	-11,13
B1	-12,42	-14,59
B2	-8,41	-6,15
C	-6,21	-
D	-50,07	-70,43

With the confidence level chosen to 95%, and the degree of freedom equal to 14 (n-1 =15-1=14) the interval in which the t-value must lie in order for the null hypothesis to be accepted equals ∓2,145. As can be seen in table 6.3, none of the recordings fulfill this requirement. The null hypothesis can be rejected for all recordings. This means that it can be said that there are significant differences between all recordings and the anechoic recordings with a 95% confidence.

The t-value of room C is closest to the interval mentioned below. Calculations show that in order for the difference not to be considered significant in this case the mean value of room C would have had to be equal to 8,96 or higher, or the value of μ_0 would have had to be lowered to a value of 8.04 or lower.

6.3. Calculation of the critical distance, *r*

The critical distance was calculated for each room configuration according to equation 2.7 and 5.2. The results are shown in table 6.4 below, room configuration A1 and A2 are shown together since they have the exact same acoustical properties and the calculation would therefore yield the same results.

Table 6.4. The critical distance for each room configuration and octave band. Mean values are shown in the bottom of the table.

Frequency [Hz] Room configuration	125	250	500	1000	2000	4000	**Mean value**
A1 & A2	1,06	1,18	1,18	1,16	1,17	1,11	**1,14**
B1	0,59	0,65	0,61	0,56	0,58	0,55	**0,59**
B2	0,55	0,75	0,90	0,80	0,59	0,55	**0,69**
C	0,99	1,33	0,98	1,65	2,17	2,69	**1,63**
D	0,28	0,32	0,29	0,32	0,37	0,40	**0,33**

As can be seen in table 6.4. all mean values of the critical distance are below 2 m, which was the distance between the loudspeaker and the artificial head. The critical distance is below 2 m for almost all cases, only in room C for frequencies of 2000 Hz and above it exceeds 2 m.

6.4. Calculation of the mean absorption coefficient, α_m

The mean absorption coefficient was calculated for each of the octave bands and for all room configurations. It is defined as the mean value of the values of the octave bands between 250 – 4000 Hz, the results are shown in table 6.5 below.

Table 6.5. The mean absorption coefficient for all octave bands and all room configurations, mean values are shown in the bottom of the table.

Frequency [Hz] Room configuration	125	250	500	1000	2000	4000	**Mean value**
A1 & A2	0,162	0,198	0,200	0,194	0,196	0,178	**0,188**
B1	0,158	0,190	0,170	0,141	0,150	0,138	**0,158**
B2	0,138	0,239	0,328	0,268	0,158	0,140	**0,212**
C	0,361	0,551	0,355	0,709	0,882	0,963	**0,637**
D	0,027	0,035	0,209	0,034	0,046	0,53	**0,037**

As table 6.5 shows, the mean absorption factor in room C in significantly higher than in the other room configurations, it is however still far from 1. In room D it is by far the lowest.

7. Discussion

7.1. Binaural recordings

How well the dimensions of the artificial head match the anatomy of the individual listener will determine how well the binaural recordings represent the original sound field. Since the same artificial head was used for all the binaural recordings, all recordings represent the original sound field equally well.

The simplified shape of the pinna of the artificial head could impair the spectral cues that is caused by the pinna. Since the task in this thesis was not to determine the position of a sound source, and the sound from the loudspeaker had frontal incidence, this did not affect the results of the listening test.

7.2. Impulse responses

The reason that room D has a much higher sound pressure level in relation to the rest of the room configurations is that since room D is a reverberation chamber, the absorption of the room is therefore very low. The room's contribution to the sound pressure level will therefore be larger.

Room D also has a significantly longer reverberation time compared to the other room configurations. This is because the low amount of absorbents present in the room. This will result in a lower amount of the sound energy being lost due to absorption each time the sound waves encounter the room boundaries. The soundwaves will continue to be reflected by the room boundaries until they are gone, since a smaller amount of energy is lost in this type of room the soundwaves can be reflected more times before they are gone. This will result in a longer reverberation time. The opposite applies for room C. The large amount of absorbents in the room absorbs a large amount of the sound energy, thus reducing the number of times a soundwave can be reflected off the room boundaries before it is gone.

As can be seen in Figure 6.2 the reverberation time for room B2 decreases between 0 and 500 Hz before it starts to increase for the higher frequencies. This appearance of the curve occurs due to flutter echo between two parallel walls in the room.

As can be seen in Figure 6.3 there is only values for the speech clarity in room configuration C for frequencies up to 1000 Hz. Speech clarity describes the quality of the speech transmission from the speaker to the listener by comparing the sound energy in the early reflections with the late reflections. Since there are a lot of absorbers present in room C there is a lack of late reflections above 1000. Because of this, the value for speech clarity in room C approaches infinity.

Room D is the only room configurations with negative values of speech clarity. This is because of the large amount of late reflections that occurs in the room due to the low amount of absorbents present.

7.3. Listening tests

As previously mentioned, the headphone-to-ear-canal transfer function has not been taken into account and no equalization has been made. Because of this the influence of the headphones on the sound will not be eliminated. Since all sound files were treated equally and the task of the listening test was to determine how big the difference between two recordings were this error has little effect on the result.

After the recordings had been done, a background noise was discovered in each of them after they had been normalized. Since the background noise appeared to be equally loud in all recordings it was assumed that it would not considerably affect the results of the listening test.

Before the listening test was conducted none of the listeners received any sort of training, they were only allowed to listen to two of the recordings each in order to familiarize with the scale used. Training of the listeners was not considered necessary because the task was not assumed to be that difficult.

As earlier mentioned, 5 of the participating listeners in the listening test had previous experience in listening tests and are therefore considered to be trained listeners. In this case it was not considered a problem using trained listeners since the task of the listening test was to detect even the smallest differences between the recordings.

By having the listeners perform the listening test more than once would have made it possible to investigate the consistency of the listeners. In this thesis each listener only performed the listening test once and it is therefore not possible to investigate their consistency.

After the test had been conducted several listeners mentioned that it had been difficult to relate to the scale during the test. They had compared previous comparisons to the ones that they were listening to in order to try to make a fair judgement. This could have affected the results of the listening test.

During the listening test the judgements of the listeners were written down on a prepared form containing a scale for each comparison they were to judge, see appendix 3. If the judgements had instead been entered directly into the computer by using a suitable software the risk for errors occurring when the data was transferred from paper to the computer could have been avoided. However, since the sample is very small in this case the risk for errors is also small. It is difficult to say if any errors occurred or how large of an impact they would have on the results in that case.

7.4. Statistical analysis

The statistical analysis of the results from the listening tests showed that there were significant differences between the binaural recordings performed in all room configurations and binaural recordings performed in an anechoic chamber. Calculations showed that if the value of μ_0 had been lowered to 8,96 or lower, the statistical analysis would not have shown a significant difference between binaural recordings performed in room C and binaural recordings performed in an anechoic chamber. Is it fair that the mean value is compared to the highest number on the scale used in the listening test in the t-test? By accepting a small difference between the binaural recordings performed in room configuration C and in the anechoic chamber the t-test could have yielded different results.

The statistical analysis shows that there are perceived differences between recordings performed in room C and recordings performed in an anechoic chamber. Figure 6.2 shows that there is a low reverberation time in room C, and Figure 6.3 shows that the speech clarity for room C approaches infinity for frequencies above 1000 HZ. This suggests that there is a lack of late reflections in room C, this means that the early reflections are important for the perceived difference between the recordings performed in the two rooms. Since there was a lot of absorbers present in room C the early reflections are possibly derived from the small details in the room, such as lighting or recording equipment that was present in the room during the recordings. This means that there could possibly be a perceived difference between recordings performed in different anechoic chambers as well, depending on which equipment that was used during the recordings. If different types of loudspeakers, artificial heads or light sources are used in different anechoic chambers, there could possibly be perceivable differences because of the differences in the early reflections.

8. Conclusion

The result of the statistical analysis of the answers from the listening test shows that there is a significant difference between how binaural recordings performed in a semi-anechoic chamber and binaural recordings performed in an anechoic chamber are perceived. To use binaural recordings performed in a semi-anechoic chamber for predicting, for instance, the sound environment in preschools by convolution with binaural impulse responses is therefore not a recommended method since it will not represent the real sound environment. The results of this thesis show that a mean absorption coefficient of 0,637 in a room isn't sufficient for there not to be a significant difference between how binaural recordings performed in the room and binaural recordings performed in an anechoic chamber are perceived. In order to achieve a larger value of the mean absorption coefficient a large amount of absorbents is needed in a room. This would make it very difficult to create this type of environment on a preschool's own premises.

The results f the impulse responses shows that there are no values for speech clarity in the semi-anechoic chamber for frequencies above 1000 Hz, this suggests that there is a lack of late reflections in the room. This means that the early reflections are important for the perceived difference between binaural recordings performed in the two rooms. This suggests that there could also be perceivable differences between binaural recordings performed in different anechoic chambers depending on the equipment used for the recordings.

9. References

Arbetsmiljöverket. (2005). *BULLER, Arbetsmiljöverkets föreskrifter om buller samt allmänna råd om tillämpningen av föreskrifterna*. Solna: Arbetsmiljöverket https://www.av.se/globalassets/filer/publikationer/foreskrifter/buller-foreskrifter-afs2005-16.pdf

Audacity (Version: 2.3.2) (2018). [Computer program] Audacity team. https://www.audacityteam.org/

Beranek, L. (2011). *The sound strength parameter g and its importance in evaluating and planning the acoustics of halls for music*. Journal of the Acoustical Society of America, 129(5), ss.3020–3026. doi: https://doi.org/10.1121/1.3573983

Bernström, R. (1987). *Akustikhandbok för arkitekter och byggnadsprojektörer*. Second ed., Stockholm: K-Konsult

Blauert, J. (2013). *The technology of binaural listening*. Berlin: Springer-Verlag

Bodén, H., Carsson, U., Glav, R., Wallin, H., Åbom, M. (2001). Ljud och vibrationer. Stockholm: KTH instutitionen fö forkostteknik

Brüel & Kjaer. (2013). *Application note, Measuring Speech Intelligibility using DIRAC Type 7841*. Naerum, Danmark: Brüel & Kjaer Sound and Vibration Measurements A/S.

David, H. (1963). *The method of paired comparisons*. London: Charles Griffin & company limited

Cohen Tenoudji, F. (2016). *Analog and Digital Signal Analysis, From Basics to Applications.* Switzerland: Springer International Publishing

Ecophon. (n.d.). *Sound strength*. https://www.ecophon.com/en/knowledge/acoustic-knowledge/room-acoustic-descriptors/sound-strength/ [2019-04-02]

Engstrand, U. & Olsson, U. (2003). *Variansanalys och försöksplanering*. Lund: Studentlitteratur

Fels, J. (2008). *From children to adults: How binaural cues and ear canal impedances grow*. Diss. Aachen: rheinisch westfälischen technischen hochschule.

Folkhälsomyndigheten. (2017). *Miljöhälsorapport 2017.* Livsmedelsverket. http://www.imm.ki.se/MHR2017.pdf?_ga=2.45546468.1730794290.1553159557-847017192.1553159557

Ginn, K.B. (1978). *Architectural acoustics.* Denmark: Brüel & Kjaer.

Hammershøi, D & Møller, H. (1992). *Fundamentals of binaural technology*. Denmark: Aalborg universitet. (Congress eller ej?)

HEAD Acoustics. (n.d.). *Application note - 07/16, Proper binaural recording and playback*. https://www.head-acoustics.com/downloads/eng/application_notes/NotesonBinauralMeasurements_e.pdf [2019-03-25]

HEAD acoustics. (n.d.). *Application note - 06/11, Conducting listening tests*. https://www.head-acoustics.com/downloads/eng/application_notes/ListeningTests_e.pdf [2019-07-01]

Howard, D., Angus, J. (2009). *Acoustics and Psychoacoustics*, fourth ed. Focal Press.

International organization of standardization. (2009). Acoustics - Measurement of room acoustic parameters - Part 1: Performance spaces – (ISO standard No. 3382-1:2009).

Johansson, A. (2005). *Drum sound from floor coverings – objective and subjective assessment*. Diss. Lund: Lund university

Kuttruff, H. (1991). *Room acoustics*. Third. ed., England: Elsevier science publishers LTD

McKenzie, T., Murphy, D, T., Kearney, G. (2019). *Interaural Level Difference Optimization of Binaural Ambisonic Rendering*. Applied sciences, (9). doi:10.3390/app9061226

Melo, V., Tenenbaum,R & Musafir,R (2013). *Intelligibility assessment in elementary school classrooms from binaural room impulse responses measured with a childlike dummy head.* Applied acoustics, 74(2013), ss. 1436-1447

Meyer, J. (2009). *Acoustics and the performance of music, Manual for acousticians, audio engineers, musicians, architects and musical instrument makers*. Fifth ed., USA: Springer
På förlag, skriva orginalutgivare eller översättare när jag använt mig av översatt version?

Moore, B. (2004). *An introduction to the psychology of hearing*. Fifth ed., Bingley: Emerald Group Publishing Limited

Nilsson, E., Johansson, A., Brunskog, J., Sjökvist, L. & Holmberg, D. (2008). *Grundläggande akustik.* 4. uppl., Lund: Lunds universitet

Resonance hearing clinic. (2013). *Ear canal resonance*. http://www.resonancehearingclinic.com/resonance-of-the-human-ear-canal/ [2019-04-02]

Rumsey, F. & McCormick, T. (2010). *Sound and recording*. sixth ed., Oxford: Elsevier LTD

Schøyen Nielsen, M. B. (n.d.). *Anechoic vs. Semi anechoic rooms*. Hørsholm Denmark: Sessions (Brüel Acoustics)
https://www.icacommission.org/Proceedings/ICA2001Rome/5_01.pdf

Sjödin, F. (2012). *Noise in preschools, health and preventive measures*. Diss. Umeå: Department of public health and clinical medicine

Socialstyrelsen. (2005). *Miljöhälsorapport 2005*. Socialstyrelsen https://ki.se/sites/default/files/mhr2005_0.pdf

Socialstyrelsen. (2008). *Buller - höga ljudnivåer och buller inomhus*. Socialstyrelsen https://www.folkhalsomyndigheten.se/contentassets/ad862888cbd54496b6aa8ec71247bd75/buller-hoga-ljudnivaer-inomhus.pdf

Socialstyrelsen (2010). *Buller bort! En liten bok om god ljudmiljö i förskolan* [broschyr]. https://www.folkhalsomyndigheten.se/contentassets/e33afaaeb0034dcabf9fd279d55f50bf/buller-bort-liten-bok-ljudniva-forskola.pdf

Söderholm, M. (1998). *Listening tests as a tool in sound quality work applied on vacuum cleaners*. Master thesis. Stockholm: KTH

Vorländer, M. (2008). *Auralization, Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality*. Berlin: Springer.

Vårdguiden. (2005). *Så fungerar kroppen, Sinnen.* https://www.1177.se/Skane/liv--halsa/sa-fungerar-kroppen/sinnen/ [2019-04-02]

Waye, K., Agge, A., Lindström, F., Hult, M. (2011). *God ljudmiljö i förskola – samband mellan ljudmiljö, hälsa och välbefinnande före och efter åtgärdsprogram*. Göteborg: Enheten för arbets- och miljömedicin.
https://gupea.ub.gu.se/bitstream/2077/35765/1/gupea_2077_35765_1.pdf

Xie, B. (2013). *Head-Related Transfer Function and Virtual Auditory Display*. Second ed., Plantation, FL: J. Ross Publishing.

Zhong, X. (2013). *Perceptual Evaluation of Inter-individual Differences in Headphone Equalization*. International conference on natural computation (ICNC). Shenyang, China, july 23-25 2013.

Zwicker, E & Fastl, H. (1999). Psychoacoustics. Second ed., Berlin: Springer Publishing

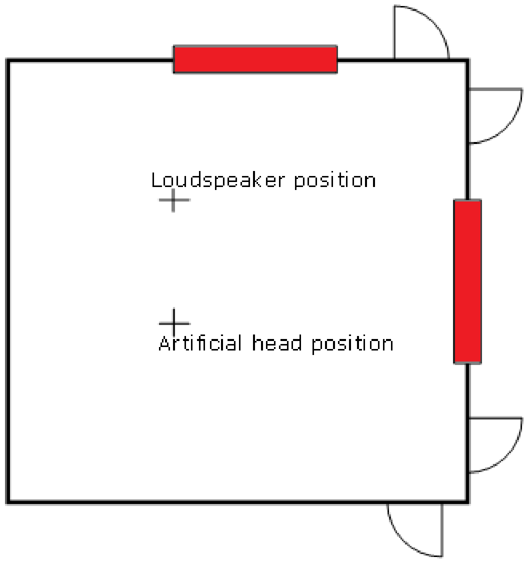
Åkerlöf, L. (2001). *Byggnadsakustik*. Stockholm: Svensk byggtjänst

10. Appendix

Appendix 1 – Floorplan sketches over for each room configuration. Placement of the loudspeaker, the artificial head and any extra insulation are shown.

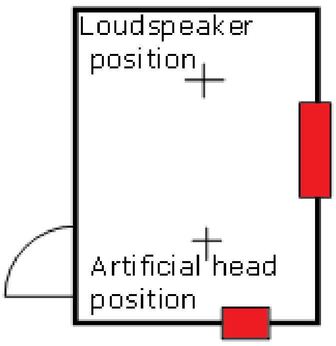
Room A

Red markings represent absorbers of the type “Wall panel A” from Ecophon.



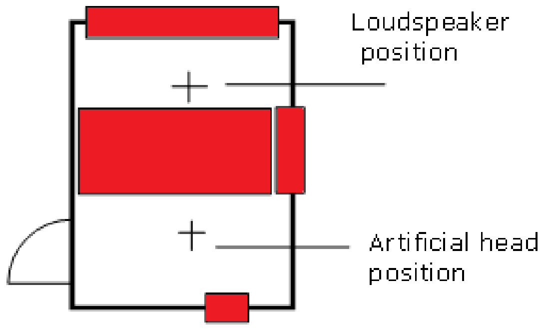
Room B

Red markings represent the two absorbers with dimensions of 1,2x1,2 m and 0,6x1,2 m present in the room.

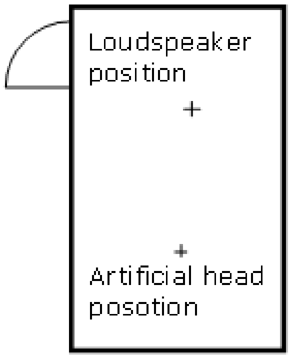


Room B + Extra insulation

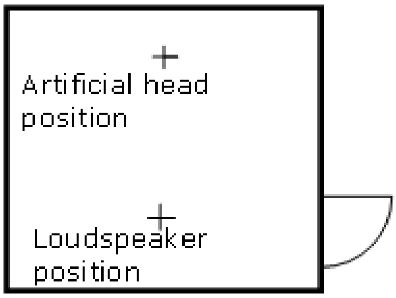
Red markings represent the two absorbers with dimensions of 1,2x1,2 m and 0,6x1,2 m present in the room with two additional absorbers of the type "Wall panel A" from Ecophon.



Room C



Room D



Appendix 2 – Questionnaire about previous experience and hearing that the listeners were asked to fill in at the time of the test. The questionnaire was given in both English and Swedish.

Questionnaire

Name:

Do you have any known problems with your hearing? ________

Do you consider yourself to have normal hearing ________

Have you participated in listening tests before? ________

Frågeformulär

Namn:

Har du några kända hörselnedsättningar? _________

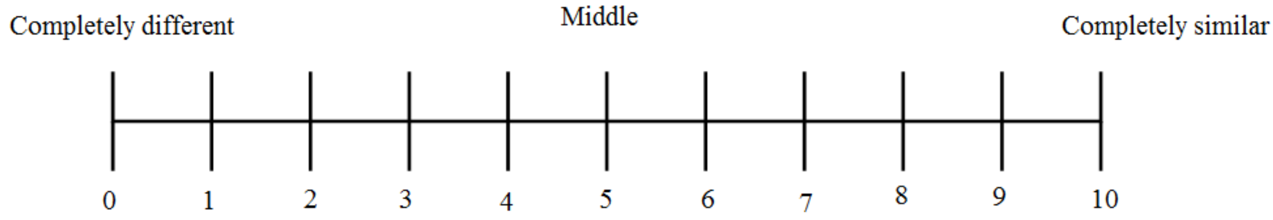
Anser du att du har normal hörsel? _________

Har du deltagit i lyssningstest tidigare? _________

Appendix 3 – Questionnaire that the listeners were asked to fill in during the listening test. The questionnaire was given in both Swedish and English.

Listening test

Hi! Thank you for participating in this listening test. This listening test is a part of my master thesis , in which binaural recordings of both male and female speech have been performed in six different room configurations. You will now listen to these recordings in pairs, for each pair you will be asked to judge how similar you perceive the recordings to be. You will be asked to answer on a scale, see figure below.



Put a mark on the scale where you think fits best with the pair of recordings you are listening to.

The recordings have been normalized so that all recordings would be perceived as equally loud. It is not the difference in loudness that are of interest to investigate in this listening test, but the differences in how the reflections are perceived. Do you hear any difference in the amount of reflections present in the two different recordings?

You will now get to listen to 11 pairs of recordings, 6 of female speech and 5 of male speech.

How similar do you perceive the two recordings to be?

- Put a mark on the scales where you think fits best with the pairs of recordings that you are listening to.

1.
Completely different — Middle — Completely similar
0 1 2 3 4 5 6 7 8 9 10

2.
Completely different — Middle — Completely similar
0 1 2 3 4 5 6 7 8 9 10

3.
Completely different — Middle — Completely similar
0 1 2 3 4 5 6 7 8 9 10

4.
Completely different — Middle — Completely similar
0 1 2 3 4 5 6 7 8 9 10

5.
Completely different — Middle — Completely similar
0 1 2 3 4 5 6 7 8 9 10

6.
Completely different — Middle — Completely similar
0 1 2 3 4 5 6 7 8 9 10

How similar do you perceive the two recordings to be?

- Put a mark on the scales where you think fits best with the pairs of recordings that you are listening to.

7.
Completely different | Middle | Completely similar
0 1 2 3 4 5 6 7 8 9 10

8.
Completely different | Middle | Completely similar
0 1 2 3 4 5 6 7 8 9 10

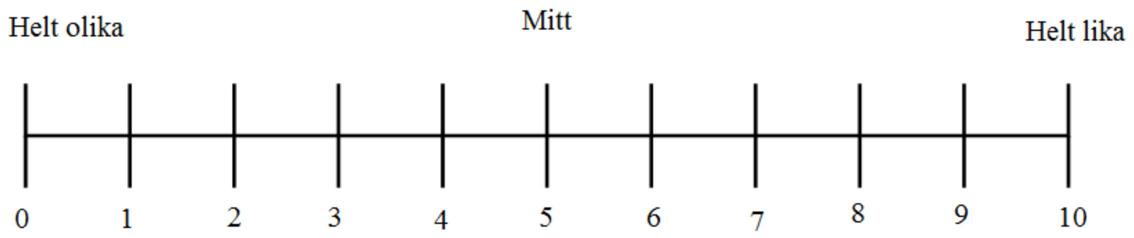
9.
Completely different | Middle | Completely similar
0 1 2 3 4 5 6 7 8 9 10

10.
Completely different | Middle | Completely similar
0 1 2 3 4 5 6 7 8 9 10

11.
Completely different | Middle | Completely similar
0 1 2 3 4 5 6 7 8 9 10

Lyssningstest

Hej! Tack för att du tagit dig tid att delta i detta lyssningstest. Detta lyssningstest är en del av mitt examensarbete där binaurala inselningar av både manligt och kvinligt tal utförts i sex olika rumskonfigurationer. Du kommer nu att få lyssna på dessa inspelningar parvis, för varje par ska du sedan svara på en skala hur lika du upplever att inspelningarna är, se figur nedan.



Markera med ett kryss eller ett streck på skalan där du upplever stämmer bäst överens med det par av inspelningar du lyssnar på.

Ljudfilerna har normaliserats för att ljudnivån i alla inspelningar ska upplevas lika högt. Det är alltså inte nivåskillnader mellan inspelningarna som är av intresse i detta lyssningstest, utan det som är intressant här är hur reflektionerna uppfattas. Hör du skillnad i hur mycket reflektioner det förekommer i de två olika inspelningarna?

Du kommer nu att få lyssna på 12 parvisa jämförelser, 6 st med kvinligt tal och 5 st med manligt tal.

Hur lika upplevs de två inspelningarna?

- Markera på skalan där du tycker att det stämmer överens med de inspelningar du lyssnar på.

Helt olika Mitt Helt lika

0 1 2 3 4 5 6 7 8 9 10

1.

Helt olika Mitt Helt lika

0 1 2 3 4 5 6 7 8 9 10

2.

Helt olika Mitt Helt lika

0 1 2 3 4 5 6 7 8 9 10

3.

Helt olika Mitt Helt lika

0 1 2 3 4 5 6 7 8 9 10

4.

Helt olika Mitt Helt lika

0 1 2 3 4 5 6 7 8 9 10

5.

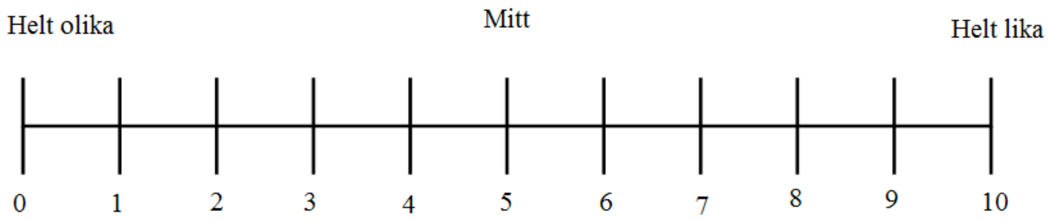
Helt olika Mitt Helt lika

0 1 2 3 4 5 6 7 8 9 10

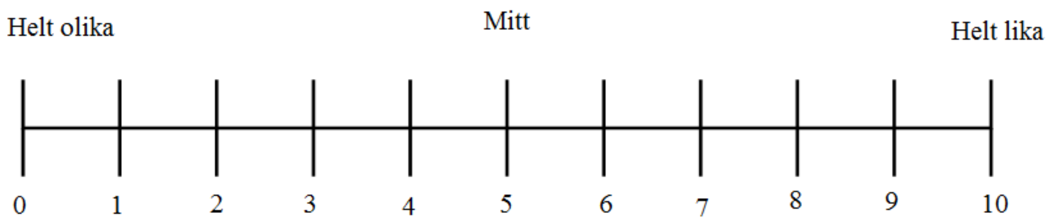
6.

Hur lika upplevs de två inspelningarna?

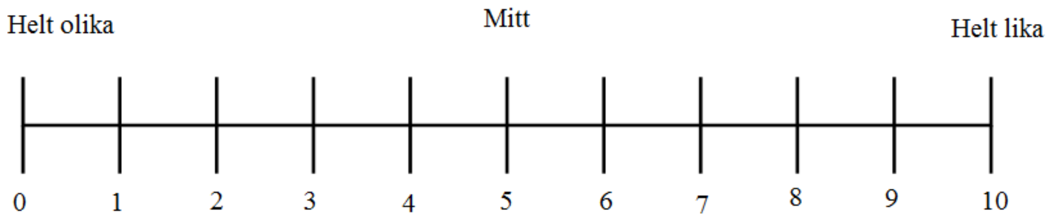
- Markera på skalan där du tycker att det stämmer överens med de inspelningar du lyssnar på.



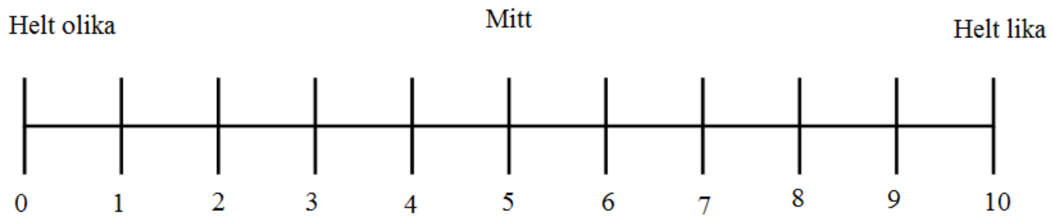
7.



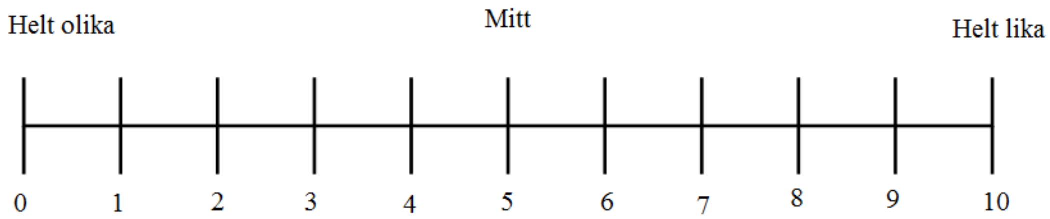
8.



9.



10.



11.