NTNU – Trondheim
Norwegian University of
Science and Technology

Source Direction Determination with Headphones

An Adaptable Model for Binaural Surround Sound

Audun Bekkos

Master of Science in Electronics
Submission date: June 2012
Supervisor: Peter Svensson, IET

Norwegian University of Science and Technology
Department of Electronics and Telecommunications

Preface

This master's thesis is the final work on my master's degree in acoustics at Norwegian University of Science and Technology. The idea behind this thesis originated in the search for a good surround sound gaming headset for personal use. After reading and viewing several reviews for different commercial products, where the same product got good and bad reviews of its surround capabilities, the idea that can be described as *" How hard can it be to create a simple model to outperform these products? "* started to grow. After doing research on the commercially used technology, use of knowledge about surround sound techniques, and a discussion with my supervisor Peter Svensson, a problem to be studied emerged. Parameters and limitations of the task were defined. It was decided that a subjective listening test should be used to evaluate the adaptable binaural models performance. Its performance would be compared to the performance of a commercial surround sound model for headphones. The adaptable model is created with different known theory on the subject of surround sound in headphones.

Thanks to my supervisor Peter Svensson for great help and supervision. Exceptional thanks to Richard Duda for sharing MATLAB files used for calculating interaural time delay for an ellipsoidal head model [1]. Special thanks to my childhood friend Mikael Berg for helping with some mathematics, and classmate Ingebjørg Nordstoga Eide for sharing information and ideas in the start up phase of this thesis. Big thanks to all the test subjects that took time to participate on the listening test, and also thanks to my room mates that pretended to be interested while I babbled on about my thesis.

Audun Bekkos

Summary

An adaptable binaural model for surround sound has been developed in this master's thesis. The adaptability is based on measurements of the listener's head. This model is based on what was found to be the best suited material combination of successful models in earlier studies. This includes an ellipsoidal model for interaural time difference, an one-pole, one-zero head shadow filter and the use of Blauert's directional bands for spectral manipulation. The model can play back six channel surround content using the standardized 5.1 surround sound loudspeaker setup. This standardized loudspeaker placement is used when creating virtual sound sources. Arbitrary sound directions are made in the horizontal plane by creating virtual sound sources using vector base amplitude panning between the standardized loudspeaker positions.

To test the performance of this model, a listening test was conducted. The hypothesis tested was that the adaptable model would produce equal or lower localization error, compared to the commercial model. 20 test subjects participated. The test featured three different test types; standardized 5.1 loudspeaker setup, a commercial model for surround sound in headphones, and the adaptable model. Localization accuracy for ten selected directions in the right half plane was tested. The results from the adaptable model were compared to the result of the commercial model. The loudspeaker setup acted as a reference.

Mean localization error was found to be thrice as high for the adaptable model, compared to the commercial model. Both models had the same standard deviation. 95% of the confidence intervals for these models did not overlap, i.e. there is a significant difference between the two methods. With this one can safely conclude that the commercial model provided a smaller localization error than the adaptable model. Hence the hypothesis has to be disproved.

Both the commercial model and the thesis model performed significantly worse than the loudspeaker setup. One difference between commercial model, and the thesis model, was that that the commercial model had added room reflections and reverberation. This can create the sensation that the sound is coming from outside the head, and make it easier to localize. This contradicts with the knowledge that reverberation diffuses the sound field, making the direct sound that provides the directional information become less prominent.

Sammendrag

En tilpassbar binaural modell for å gjengi surroundlyd har blitt utviklet i denne masteroppgaven. Tilpassbarheten er basert på målinger av lytterens hode. Modellen er satt sammen av et utvalg vellykkede modeller fra tidligere arbeid som passet seg best til formålet. Dette inkluderer en ellipsiode-modell for beregning av interaural tidsforskjell, et enkelt filter for hodeskygge, samt bruk av Blauerts retningsbånd til manipulasjon av frekvensspekteret. Modellen kan spille av sekskanals surrondlyd materiale som bruker det standardiserte 5.1 høyttaleroppsettet. Denne høyttalerplasseringen brukes til å lage virtuelle kilder. Vilkårlige lydretninger, ved å lage virtuelle kilder, kan gjengis i horisontalplanet ved bruk av vektorbasert amplitudepanorering mellom disse høyttalerposisjonene.

En lyttetest ble gjennomført for å teste ytelsen til modellen. Hypotesen som skulle testes var at denne tilpassbare modellen vil gi lik eller lavere lokaliseringsfeil, i forhold til en kommersiell modell. 20 testpersoner deltok. Testen besto av tre forskjellige testtyper: det standardiserte 5.1 høyttaleroppsettet, en kommersiell modell for surroundlyd i hodetelefoner, og denne tilpassbare modellen. Lokaliseringsnøyaktighet for ti valgte retninger i det høyere halvplan ble testet. Resultatene fra den tilpassbare modellen ble sammenlignet med resultatene til den kommersielle modellen. Høyttaleroppsettet ble brukt som en referanse.

Middelverdien av lokaliseringsfeilene til den tilpassbare modellen viste seg å være tre ganger så stor, sammenlignet med den kommersielle modellen. Standardavviket var likt for begge modellene. 95% av konfidensintervallene for de to modellene overlappet ikke, det vil si at det er en betydelig forskjell mellom de to metodene. Med dette kan man trygt konkludere med at den kommersielle modellen gir mindre feil enn den tilpassbare modellen. Følgelig kan hypotesen forkastes.

Både den kommersielle modellen og den tilpassbare modellen ga betydelig dårligere resultater enn høyttaleroppsettet. En forskjell mellom den kommersielle modellen, og den tilpassbare modellen, var at at den kommersielle modellen bruker romrefleksjoner og etterklang. Dette kan skape en følelse av at lyden kommer fra utenfor hodet, og gjøre det lettere å lokalisere. Dette står i strid med kunnskapen om at etterklang gjør lydfeltet diffust, noe som gjør at direkte lyden, som inneholder retningsinformasjonen, blir mindre fremtredende.

Contents

Chapter 1

Introduction

1.1 Background and motivation

The idea and concept of recording or synthesising directional sound, and presenting it over headphones or loudspeakers to recreate the directional sense of the sound field, has existed since stereo sound was invented. Stereo imagining works great for both loudspeakers and headphones, giving a high resolutions sound stage between the two speakers, or in headphones, far right to far left. The first occurrence of surround sound, as we know it today, was in 1940 [2]. Walt Disney's *Fantasia*, which featured music sections by Stokowski, used multiple horn loudspeakers which was fed a manually manipulated signal to create the effect of a moving sound source. Since then surround sound has become the standard format for most films that are shown in theatres, and later sold on DVDs and Blue-rays. This surround format requires a given amount of loudspeakers placed in a certain way. The loudness, price or area occupied by loudspeakers are not always convenient. Both the great masses and engineers saw the need for the possibility to playback of surround content in headphones. Since stereo in headphones worked out great, why not surround sound in headphones? In the later years the availability and popularity of surround headphones, or headset[1], has exploded on the commercial marked. These surround headphones can be divided into two main categories; headphones that use the placement of multiple loudspeaker elements to create the surround effect, and headphones that use binaural techniques (section 2.3) and signal processing to create the surround effect. The latter kind is most common. Both types have the weakness that their method of creating surround sound is static, relative to the listener: either signal processing based on a static model, or static placement of the elements in the headphone. This results in that people with physical attributes that differ from the one used in the surround sound model, will not achieve the intended directional effect. This is a big design flaw, since most of us have different head shapes, or atleast different pinna shapes or ear placement than in the model used. The best result could be achieved by measuring the head related transfer function (HRTF) for each listener, and for each source direction. More on HRTFs in section 2.1.4. This requires special and expensive equipment, and anechoic environment. It is also time consuming and somewhat cumbersome to measure, both for the supplier and the buyer. This service would also be expensive to buy, which will limit the marked drastically. This is why most manufacturers do not supply this. Instead they land on a middle ground solution that will work good for some people, but will just give some sense of direction for the majority. This is sufficient

[1] Headphone or headset? The main difference is that headsets comes with an attached microphone, while headphones are just headphones. Headsets were commonly known as something used for communication purposes, and often offered poorly sound quality compared to headphones. In the later years, gaming headsets combine high quality headphones with a microphone. The sound quality gap between headphones and headsets is closing.

for many applications, such as films. The exact direction of where the sound comes from is not critical to neither the film viewing experience, nor the plot of the film. Films are also produced in such a way that the important events are shown on the screen. This limits the span of directions to between the front loudspeakers, and hence stereo panning is sufficient. The surround loudspeakers are mainly used for ambient sounds to give the sensation of being within the sound field. Another application for surround sound is video games. These depend much more on accurate representation of the sound direction. If the player can have the visual focus on one thing, but also be aware of what is happening behind him or above him at he same time, it will give the player a great advantage and a richer gaming experience. These two applications are for the great masses. Applications such as realistic simulators for soldiers, ships, aircraft, fighter planes, or the communication system of air traffic controllers, also depend on accurate directional representation sound. These applications are for a smaller audience. A more expensive model, such as measurement of individual HRTF's, might be preferred. In spite of the bad description of the commercial used models above, they sell really well. Is this because there is nothing better to buy? Do they supply the sufficient surround experiment, even with their flaws? Or is it just that it is really difficult to create something which is better than what is already on the marked? If an adaptable accurate surround sound model for headphones were created, and it outperformed the existing models, the marked would certainly welcome it.

1.2 Overview and limitations

There has been done extensive research and tests on this subject in the past, and will surely continue being a subject of great interest in the future. The result might be that there will never be an adaptable theoretical model that can create exact directional surround sound for everyone. This task can be compared to replicating the exact fingerprint of a person, since the shape of e.g. the pinna is equally individual for every person. However, what is "exact"? In our digital age, even the best quantization introduce an irreversible error. Hence the exactness, if achieved, is gone. The goal should be to create a model where the directional error is below a given value, for a large amount of the population. The need for an exact solution might be overrated in compared to gains of such a model. Why use tremendous amounts of time, money and effort creating something that we might not even hear, or be able to differentiate? The need of pinpointing some foe behind you down to a single spatial degree in a video game is not necessary. A precision of 45, 20, 10 or 5 degrees can be sufficient for this application. Some research makes use of their own designed stimuli to test their models, e.g. virtual sources generated with a single degree accuracy in both azimuth (horizontal spatialization) and elevation (vertical spatalization). This kind of stimuli is rarely found outside the lab. The most commercial available surround content today are DVDs, Blue-rays and video games. The most common surround playback system for these applications are a 5.1[2] or 7.1 loudspeaker setup. Other surround sound loudspeaker setups worth mentioning are Ambisonics and wave field synthesis (WFS). These two setups will be shortly presented in the theory sections 2.2.3 and 2.2.4. The 5.1 setup is the most common way to playback commercially available surround sound, and the amount of available content which fits this playback setup is pretty much unlimited. This is why this thesis will be based on implementing a model that can make use of this standardized content. This is chosen to make the research and results found in this thesis more transferable to every day use. Both 5.1 and 7.1 setups positions their five or seven loudspeakers in the horizontal plane at ear level. This makes them poorly suited to playback elevated sources. For these reasons, the focus will be on implementing and testing for virtual sources in the horizontal

[2] X.1 refers to a loudspeaker setup with X loudspeakers around the listening position. The .1 refers to a low frequency channel for low frequency effects (LFE). The .1 is usually implemented with a subwoofer.

plane in this thesis. This testing will compare two models; one commercially available, and one based on the theoretical results from different articles and studies. There are multiple commercially available models for surround sound in headphones. Dolby Headphones™ have been chosen for this thesis. The intended 5.1 surround loudspeaker setup will act as the reference. More details about this test setup can be found in the experiment chapter 4. The goal of this master's thesis is to test the following hypothesis.

Hypothesis:

> The adaptable model will perform as well, or better, than the commercial model in localization virtual sound sources. In other words, the adaptable model will give a smaller, or equal, error compared to the error of the commercial model.

1.3 Structure and method

Following this introduction there is a presentation of the theory used in this thesis, chapter 2. This chapter will include some general theory on how we humans can differentiate directional incoming sounds. Different binaural syntheses and surround techniques will be presented. All the theory used in this thesis is well researched and well established. The implementation, chapter 3, will describe how the adaptable model was created: what was implemented and why. A listening test was used to test the model. The details around this listening test are described in the experiment chapter, chapter 4. Results are then presented and discussed. The hypothesis is revisited in the conclusion, chapter 7, and either approved or disproved. The last chapter will present some thoughts on which direction the future of surround sound should take. Additionally, the use of *we*, *our* or *us* in this thesis refers to us humans.

Chapter 2

Theory

This chapter will present the theory used in this thesis. First a short overview on how directional hearing works. The basics of binaural synthesis will be presented, followed by some different surround techniques. Challenges and pitfalls of binaural synthesis will be briefly mentioned. The last section in this theory chapter will look into how one can evaluate binaural synthesis and directional hearing models.

2.1 Directional hearing

The fact that we humans, and also most animals, can differentiate sound from different directions is no new knowledge. We use it all day, every day, and without thinking about it. It just works. But what makes it work, and can it be replicated? Our directional hearing mechanism can be divided into three main components; interaural time difference (ITD), interaural level difference (ILD) and spectral cues[1].

2.1.1 ITD

The interaural time difference comes from the fact that the speed of sound has a finite value, combined with the fact that our ears are placed some distance apart. ITD is the difference in distance from a source to the closest ear, and the other ear, divided by the speed of sound. This arrival time difference will vary with azimuth and elevation. There is not a one to one relationship between the azimuth and elevation combination and the ITD. Multiple azimuth and elevation combinations can give the same ITD. This effect is known as the cone of confusion. More on this effect in the challenges and pitfalls section later. The ITD is somewhat unique for a given individual. It is related to the size of the head, and the placement of the ears. A given ITD for one person will not, in most cases, give the same azimuth and elevation impression for another person. This leads to that the use of a fixed ITD model will give errors for most of the population. The simplest model for ITD is for a spherical head model with the ears placed at ±90 degrees. The ITD in the horizontal plane for this model can easily be calculated [3] with the following formula 2.1. Where a is the head radii and θ is the angle between the source location and the median plane. This formula assumes that the source is infinite distance away.

$$ITD = \frac{a(\theta + \sin(\theta))}{c} \tag{2.1}$$

The value of ITD typically varies in range of 0 – 800 µs. The ITD is frequency independent below approximately 500 Hz, and above 3000 Hz. The ITD below 500 Hz can be said to be $\frac{3}{2}$ of the ITD value above 3000 Hz. In between these two frequencies, the ITD is dependent on

[1] Spectral cues are distinctive, personal, peaks and notches in the frequency response of our hearing.

frequency, and has its minimum values between 1400 and 1600 kHz [4]. ITD is the dominant mechanism for locating sound sources below 900 Hz. Between 900 and 1250 Hz there is an ambiguity. In this frequency range the phase of the incident sound can correspond to two different incoming angles, depending on which ear can be said to be the leading one. Above 1600 Hz, our hearing mechanism cannot follow the variation in the signal, and the ITD is based on the envelope delay between the two ears. From about 1250 Hz the effect of interaural level difference becomes more significant.

2.1.2 ILD

The interaural level difference occurs because the head consists of a denser and more rigid material than air, and because the head has some spatial dimension. The ILD varies with frequency. k denotes the wave number, and a is still the head radii. For frequency where $ka \ll 1$, the head does not affect the sound pressure distribution around the head significantly, and the pressure is equal for both ears regardless of sound incident angle. For higher frequencies where $ka \gg 1$, the head will be of considerable size compared to the wavelength. This creates an obstacle for the sound wave. Some part of the sound wave will get reflected off the head, and some will diffract around the head. This gives a pressure build-up on the side facing the sound source, and a lack of pressure on the shadow side of the head. As mentioned in the subsection above, this is the most significant mechanism for locating sound sources in the high frequency range.

2.1.3 Spectral cues

The head and body can be said to be pretty symmetric down the middle, creating the median plane. The median plane consists of what is right in front of us, right above us and right behind us. The ITD and ILD is equal for both ears when the sound source is in this plane, which renders the two mechanisms useless. This is where spectral cues kick in. Spectral cues can also be called monaural cues, since this mechanism also works if one is only able to listen with one ear. ITD and ILD are binaural, i.e. require the use of both ears. Spectral cues are created by reflection from the shoulders and torso, as well as the more important reflections and resonances in the pinna. The shape of the pinna can be said to be equally individual to a person as their fingerprint. It is inferred that the daily experience with real sources, combined with a subconscious learning process, have created a kind of individualized spectral cue database. This database has been constantly updated while we grow up and change both head, pinna and torso size. It also works if we wear hats, have different hairstyles or wear big jackets. It will over time even adapt to the loss or alterations of the outer ear or torso. There is for example a common plastic surgery used on children to reduce protruding ears, so they don't get bullied at school. Torso alteration can be losing an arm at the shoulder joint, or just simple muscle or fat gain/loss. In other words, the spectral cue " database " can be said to be quite robust, but it is still very much individualized. Cues not familiar to us, or familiar cues for another person, will not trigger this mechanism and give a correct localization of the sound source.

2.1.4 HRTF/HRIR

Head related impulse response (HRIR) describes how an impulse sound played at a given position, will have been altered by our head, torso and pinna before it reaches the ear canal. HRIR is a time function. Head related transfer function (HRFT) is the frequency-domain version of the HRIR. Both HRTF and HRIR come in unique pairs, since we humans have two ears. These pairs are also unique for the given source position and acoustical environment. It will also be affected by the source i.e. a loudspeaker. This can be compensated for by using an inverse filter of the source's frequency response, to only get the transfer function that describes how our ears and body filter the sound from a given direction. Anechoic environment and highly dynamic, and neutral, loudspeakers and in-ear microphones are usually used for recording HRTFs. HRTFs combine the three attributes mentioned above; ITD, ILD and spectral cues in one simple FIR[2] filter, which can easily be implemented. HRIR can be used for binaural synthesis by convolving the HRIRs with an arbitrary sound. The result played back over headphones will create the effect that the sound is played from the direction the HRIRs was measured at. This effect is close to reality when played back with the persons own HRIRs. For everyone else, there will be some effect, but not the intended effect.

As mentioned, the measurement procedure for HRTFs is quite cumbersome and time consuming. Even the slightest movement of the head can create deviations from the results wanted. This is why researches have created models, either mathematically or physical, to use instead of a human test subject. These models vary in complexity. The simplest form is a time delay representing the ITD. This may be improved by adding a simple filter that increases or decreases parts of the frequency spectrum to emulate head shadow. Both spherical and ellipsoidal head models are found implemented both as mathematics, and as physical models with microphones simulating the ears [1, 5, 6]. The most realistic model is called a dummy head. These are used for most commercial applications and research purposes. A dummy head is a molded version of a generalized human head. It has the average shape and attributes of a head, and often some sort of natural looking ears. It can also have a torso. These dummy heads are the perfect measurement subjects, since they can sit perfectly still, and do so for hours, even days without complaining. The problem is that with all its generalization, the result will also be accordingly general. Even though the result is not of the best sort, it is still widely used because it is so practical and cheap, compared to the time and effort of measuring on all the user of the given application.

A popular dummy head is the KEMAR (Knowles Electronics Manikin for Acoustic Research). HRIR pairs for 710 different directions measured on a KEMAR became publicly available [7] by the MIT (Massachusetts Institute of Technology) in 1994. Figure 2.1 and 2.2 show a HRIR and HRTF pair from this database. Plots include a comparison with a perfect impulse, which result in a perfectly flat frequency response. From figure 2.1 one can see the obvious ITD between right and left ear. From figure 2.2 one can see the main ILD between right and left ear. With a slightly higher pressure build-up at the right ear (facing the source) and a high frequency pressure drop at the left ear (in the shadow of the head). Peaks and notches are spectral cues created by pinna shape and torso reflections. The ultimate design frequency response of a loudspeaker would be as flat as the first plot in figure 2.2, so it would play all frequencies equally.

[2] FIR (Finite Impulse Response) is a digital filter consisting of a combination of amplifications and delays. The response is finite, hence FIR, and will eventually die out to the value zero, compared to IIR (Infinite Impulse Response) which have an internal feedback so the filter really never dies out, even though it can converge to zero, but never reach zero.

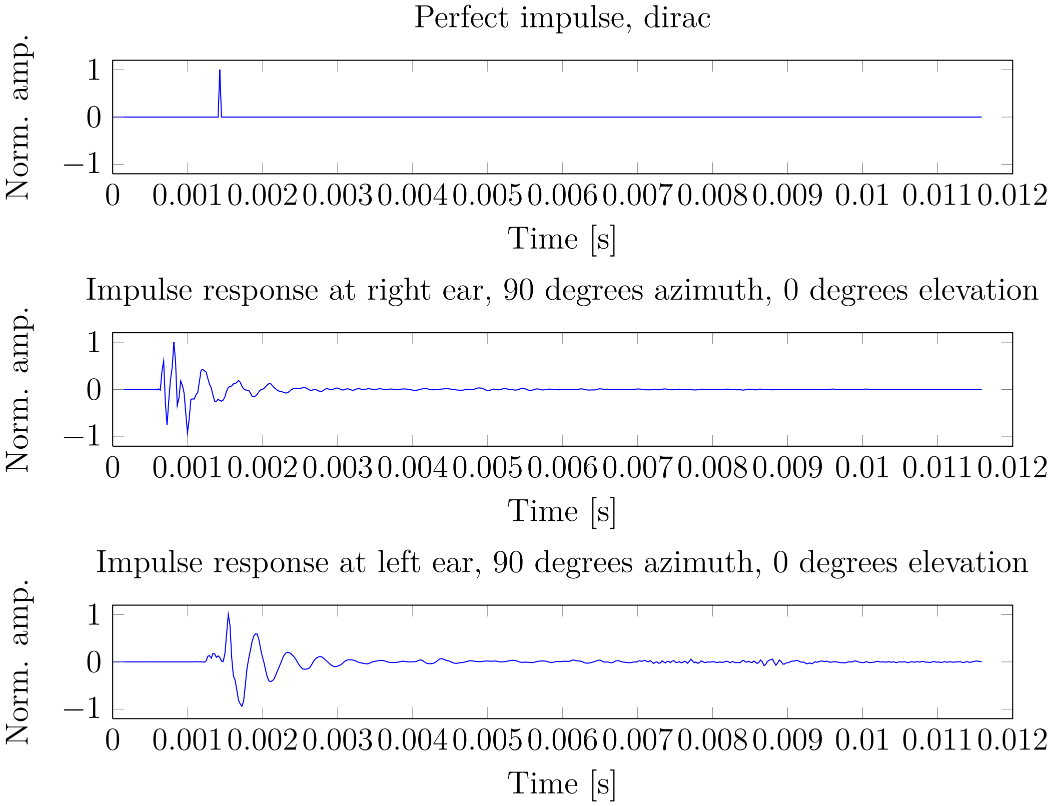


Figure 2.1: Plot of impulse responses corresponding to perfect impulse and HRTF pair 90 degrees azimuth, 0 degrees elevation. HRTF pair from the MIT database [7].

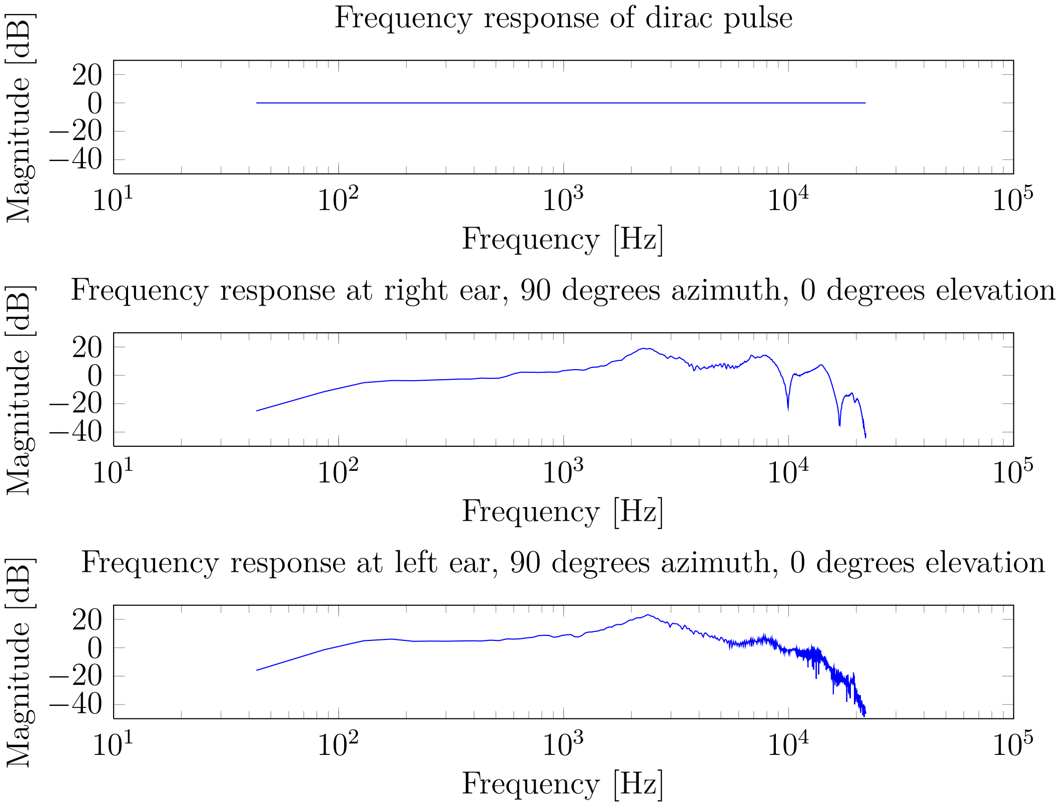


Figure 2.2: Plot of frequency responses of the impulse responses in figure 2.1. The effect of the loudspeaker used for the measurement have not been eliminated by inverse filtering for this illustration. This gives the low frequency roll-off effect seen in the figures.

2.2 Different surround sound loudspeaker setups and techniques

This section will present some of the most known surround sound loudspeaker setups, as well as the techniques used.

2.2.1 Vector Base Amplitude Panning (VBAP)

Ordinary intensity panning divides the amplitude linearly between the two loudspeakers one wants to create the virtual source between. The virtual source will then appear on a straight line between the two loudspeakers. If one sits at an equal distance from both loudspeakers, then the virtual source generated in the middle will appear closer than the loudspeakers. The reason for this is that the intensity at the listening position, when both loudspeakers play half the amplitude, will be $0.5^2 + 0.5^2 = 0.5$. This is due to that sound pressure level (SPL) is proportional with the amplitude squared. To compensate for this, one can use VBAP. VBAP creates an arc between the two loudspeakers where the virtual sources appear. The arc is the same as a thought circle passing through both loudspeakers. Figure 2.3 illustrates the placement of virtual sources with VBAP.

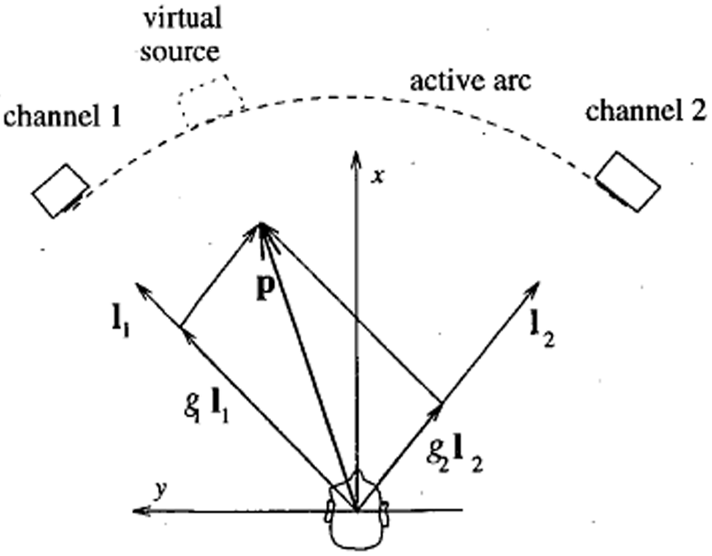


Figure 2.3: Illustration of arc for virtual sources between two loudspeakers using VBAP. Illustration taken from [8].

VBAP is not limited to only two loudspeakers, but can be used on any number of loudspeakers, which can be placed arbitrarily around the listening position. VBAP is applicable in both 2D and 3D. The general case might look something like figure 2.4. Where the vectors $\mathbf{l_1} = [l_{11}\ l_{12}\ l_{13}]^T$, $\mathbf{l_2}$ and $\mathbf{l_3}$ are unit vectors describing the direction of loudspeaker 1, 2 and 3, with reference to the center listening position. If the vector $\mathbf{p}$ points to the position where one wants the virtual source to appear, then the different gains for the three different loudspeakers can be calculated with equation 2.2.

$$\mathbf{g} = \mathbf{p^T L_{123}^{-1}} = [p_1 p_2 p_3] \begin{bmatrix} l_{11} & l_{12} & l_{13} \\ l_{21} & l_{22} & l_{23} \\ l_{31} & l_{32} & l_{33} \end{bmatrix}^{-1} \tag{2.2}$$

The gain-matrix $\mathbf{g}$ should then be scaled before used. as shown in equation 2.3, where C is the constant amount of amplitude one want to distribute.

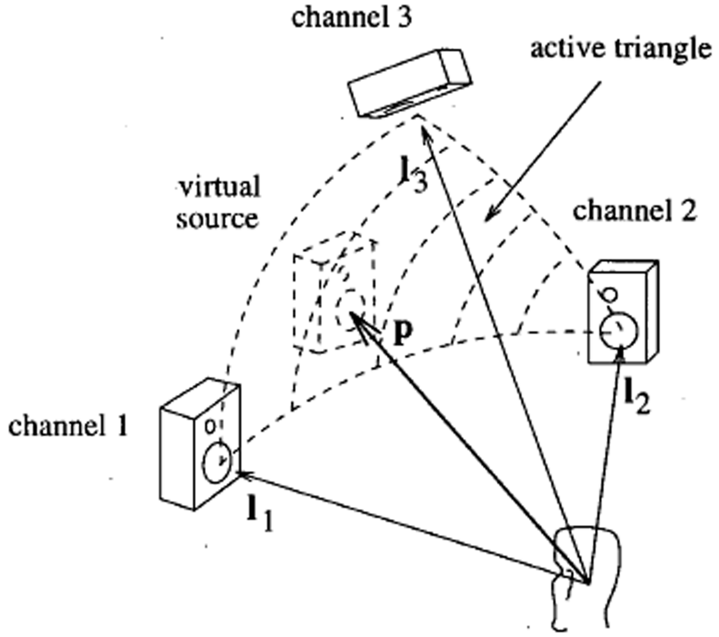


Figure 2.4: Illustration of general 3D VBAP. Taken from [8].

$$\mathbf{g}_{\mathbf{scaled}} = \frac{\sqrt{C}\mathbf{g}}{\sqrt{g_1^2 + g_2^2 + g_3^2}} \tag{2.3}$$

VBAP is used for creating virtual sources for loudspeaker setups such as stereo (2.0), quadraphonic (4.0), 5.1 and 7.1.

2.2.2 5.1 Surround Sound

As mentioned earlier, the 5.1 loudspeaker setup is the most commonly used surround sound setup, and has virtually unlimited standardized material commercially available. This setup is described in the ITU-R recommendation BS.775-2 [9]. The recommendation suggests, among other things, the optimal the loudspeaker positioning. Optimal speaker placement for 5.1 is shown in figure 2.5. Dolby has its own version of the speaker placement guide [10]. The two main sound format suppliers used with this loudspeaker setup are Dolby Digital and DTS (Digital Theatre Systems).

2.2.3 Ambisonics

7.1, 5.1 and 4.0 loudspeaker setups only aim to create the illusion of surround sound. While Ambisonics [11] aims to recreate the physical sound field recorded at the microphone position. Ambisonics needs at least four loudspeakers for playback of its basic version; *"first-order Ambisonics"*. First-order consists of four audio channels; W, X, Y and Z. W is the mono sound, and X, Y and Z contain the directional information. This format is called the B-format, and is based on a spherical harmonic decomposition of the sound field. These directions are created by using a three figure of eight microphone, or a Soundfield microphone[3], to do the recording. These four channels are then decoded into the loudspeaker signals, which is just a linear

[3] The Soundfield microphone uses four microphones placed in a tetrahedron. This microphone can be used to mono, stereo as well as surround recordings, and is frequently used for Ambisonics. The Soundfield microphone outputs the B-format after some internal processing.

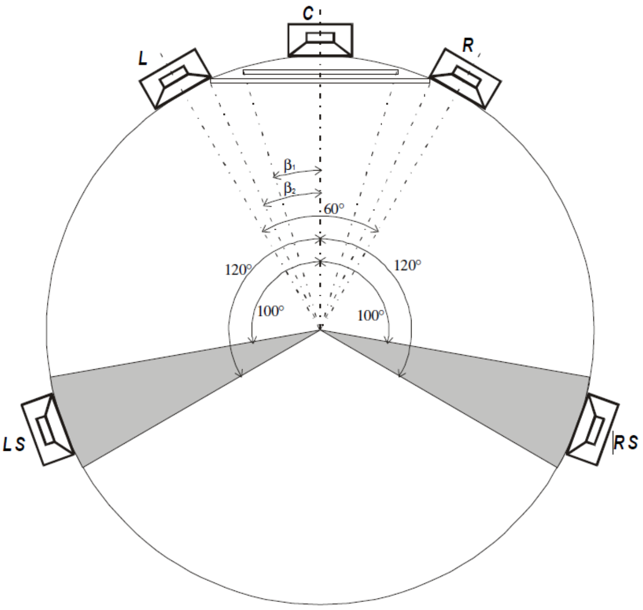


Figure 2.5: Illustration of the suggested loudspeaker setup taken from [9].

combination of the four channels W, X, Y and Z. This depends on the placement of the loudspeakers. Instead of only using two or three loudspeakers to create a virtual source as in VBAP, all loudspeakers are used. The more loudspeakers, the higher order Ambisonics possible, and a better approximation to the recorded sound field. Ambisonics have not yet made it as a commercial success, and the available material is very limited. There have also been attempts to use Ambisonics to create surround sound in headphones [12].

2.2.4 Wave Field Synthesis

Wave Field Synthesis (WFS) [13] also aim to reproduce the recorded sound field, not only at the microphone position, but over an enclosed volume. WFS is based on the Huygens principle; a wave ront can be thought to consist of a number of point sources radiating spherical waves. The superposition of all these wavefronts creates the original wavefront. This principle is used to explain, among other things, diffraction. Linear acoustic theory states that an arbitrary sound field can be created in an enclosed volume by a distribution of monopole and dipole sources placed on the surface of the volume. This has been derived mathematically in the Kirchhoff-Helmholtz integral. In a practical sense, by placing a large number of relatively small loudspeakers close to each other around the enclosed volume, and feed them with the correct signal relative to each other, one can recreate any sound field within the volume. This works up to the frequency where the wavelength starts to match the distance between the loudspeakers. One can also record the wave field one wants to recreate. The most common procedure is to use equally many microphones as loudspeakers, placed at the same relative location as one will be placing the loudspeakers. WFS requires large numbers of loudspeakers and customized content, and its use so far is therefore limited. There exists some WFS installations in cinemas, concert venues and theme parks around the world.

2.3 Binaural synthesis

This section will present the evolution of binaural synthesis, but in fairly coarse steps. From the most basic creation of an virtual sound source, to multiple virtual sources that interact in a virtual acoustic environment.

2.3.1 One virtual source

The simplest binaural synthesis is to create one virtual source on the left-right axis, either with headphones or loudspeakers. It requires two loudspeakers placed some distance apart symmetrically in front of you, or headphones being able to play different signals for each ear. By adding either time delay or amplitude difference between the two signals, the sense of a virtual source appears. For loudspeakers this source will appear somewhere between the two loudspeakers. For headphones it can vary between far left, to far right. The time delay should be in the range of 0 – 800 µs. Larger delays will just place the source at one of the loudspeakers, or far right or left in headphones. Even larger delays will create echoes, which is unwanted in this case. This thesis will focus on reproduction over headphones. The single virtual source does not need to be limited to only appear on the left-right axis. By filtering the sound with a HRTF pair, one can place the virtual source at any direction. The HRTF pair needs of course to be recorded from the same direction. One virtual source at the time makes for limited applications, but the concept is important. Next subsection will describe how this simple concept can create whole worlds of sound.

2.3.2 Superposition and auralization

In most everyday situations, there is more than one source emitting sound. Just when writing this, there is birdsong and traffic outside my window, my computer is humming, the keys on my keyboard are chattering as I type, my roommate is listening to music down the hall and my mouth is chewing gum. All these sounds exists at once, and bounce around in the room I am sitting in. By the nature of soundwaves they all end up in my ear canals. Even though they are all mixed together, I can differentiate between them, and tell which direction they are coming from.

This sound environment could be created artificially, and played back over headphones to give the same sound impression. This adding of different sounds can be done with signal processing. It is done before the sound is sent into the ear canal, but the same directional information and soundstage is preserved. This is all possible due to superposition. Superposition can be described in short that if one has two or more signals that are uncorrelated with each other, one can add them arithmetically together to create a new signal. In the scenario above, an example of superposition as valid can be that my typing frequency does not follow the rhythm of the birdsong outside the window, or that the traffic does not only make noise when I chew on the chewing gum.

Creating real-like or fake sound environments is called auralization[4], and it makes great use of this superposition principle. Auralization such as listening to how a designed building construction will sound before it is build, or added effects to video games to give the gamer the sound stage of a given environment, are widely used. Imagine a game character in an open mountain environment. The typical sound stage of such an environment is distinct, but somewhat faded, echoes from the nearby mountains. The game developer does not have to actually create a model of these mountains, and let the sound waves travel to the mountain and back, to get the wanted sound effect. The developer can easily just add a time delayed copy of the original signal, with a delay matching the double distance to the mountain. By reducing the amplitude with 1/distance, and maybe add some sort of low frequency filter effect, the illusion that this sound has travelled to the mountain and back is created.

[4] Auralization is the technique of creation and reproduction of sound on the basis of computer data.

Superposition appears naturally in our ears when listening to stereo loudspeakers. Sound from both loudspeakers reach both ears. This does not happen with headphones. Bauer found a solution to do this over headphones in 1961, which will be presented in the next subsection.

2.3.3 Stereo loudspeaker experience over headphones

A stereophonic signal played over loudspeakers will, as mentioned, create virtual sources between the two loudspeakers. This is what the producer intended. When the same signal is played over headphones, the virtual sources will vary between far left, and far right. The spatial perspective of the soundstage will appear distorted, in comparison with playback over loudspeakers. The reason for this is that the signal from the left loudspeaker reaches both the left and right ear, even if the right loudspeaker is not playing, and visa versa. That crosstalk effect is missing when listening over headphones. The lack of crosstalk creates an unnatural sound experience, which is very rarely found in nature. By introducing crosstalk between the two headphone channels, one can resurrect the intended space perspective, and create the loudspeaker listening experience. Bauer suggested such a crosstalk introduction system back in 1961 [14]. It was an analogue electrical circuit using passive circuit elements. Today this can easily be done digitally, using software to control a micro controller or DSP[5]. This Bauer circuit live today as modernized version, and goes by the name Bauer stereophonic-to-binaural DSP, and can be found at [15].

2.3.4 Surround loudspeaker setup in an acoustic environment over headphones

Bauer's circuit creates two virtual loudspeakers placed in front of the listener. There is nothing limiting these virtual speakers to only be placed in the front of the listener. By making only small changes in the crosstalk introduction filters, they can appear to be placed at the side, or even behind the listener, or any other positions for that matter. The number of simulated loudspeakers are not necessarily limited to only two either. With the power of the superposition principle, the potential of simulated loudspeakers is virtually infinite. Products offering multiple virtual speakers, combined with the auralization of virtual rooms designed to enhance the listening experience, have become more and more common in later years. There are many different types of software, patents and different approaches to recreate the surround sound loudspeaker experience over headphones. One of the most known is Dolby Headphone™ (DH). The DH technology was originally created by the Australian company Lake Technologies (also known as Lake DSP), but later sold to Dolby Laboratories in 1998. DH can be found as a software plug-in for software such as Power DVD, or hardware implemented with its own DSP in headsets, laptops, sound cards, A/V receivers and even cellphones. It will work with all types of headphones, as long as the playing device has DH. All these products are marked with the distinctive Dolby Headphone™ logo shown in figure 2.6.

Figure 2.6: Dolby Headphone™ logo

DH recreates the same loudspeaker setup suggested for the Dolby Digital 5.1 Surround Sound [10], as illustrated in figure 2.5. DH comes with three different acoustics environments; studio,

[5] Digital Signal Processor: A processor optimized to do signal processing operations. This is typical multiplications and additions. "Ordinary" processors are more all around and can handle logical expressions better.

cinema and hall. What these changes exactly do, is not publicly available, but the imminent effect is a change in reverberation and feeling of space. DH can play 5.1 encoded sound, such as the sound on DVDs and Blue-rays, or in video games. DH can also upmix standard stereo sound to surround sound. The exact technology behind DH is a company secret of Dolby. The only clue can be found in the presentation made by Adam McKeag and David S. McGrath from Lake DSP; *Using Auralisation Techniques to Render 5.1 Surround To Binaural and Transaural Playback* [16]. They presented this at the 102nd Audio Engineer Society convention in 1997. This is probably a description of what later became Dolby Headphone™ after Dolby bought it. The AES-presentation is very vague, and does not go into details in how it is implemented. It only suggests how it can be done. The presentation is honestly more of a commercial sales pitch than a presentation of technology. Dolby states the following on their website when it comes to how DH works. [17–19].

> *It's an astonishing acoustic illusion, and here's how it works: when you listen to entertainment through your stereo or home theatre system's speakers, your ears receive sound directly from each speaker and from multiple reflections of the sound on the room's surfaces and furnishings.*
> *Using powerful digital signal processing technology, Dolby Headphone recreates this audio experience, producing the sonic signature of a speaker system properly placed in a carefully defined acoustic environment. The result is a spacious audio landscape with you at the center.*
>
> *Dolby Headphone electronically imparts the sonic signature of a corresponding speaker properly placed in a carefully defined acoustic environment to each audio channel (two on stereo programs, and up to five on surround programs). The subwoofer signal (the ".1" Low-Frequency Effects [LFE] channel) is mixed into the Left and Right channels in equal proportion.*
>
> *Dolby Headphone technology is universal to all listeners due to a significant breakthrough in signal processing technology. The system does not require custom head-related transfer function (HRTF) settings to accommodate the differences between individuals, so it's simple to implement and operate.*

2.3.5 Challenges with binaural synthesis

This subsection will mention some common challenges and pitfalls with directional sound over headphones.

Front-back confusion

One of the most common problems with sound over headphones is the difficulty of differentiating between frontal sounds and sounds from behind. This phenomena is also well-know as the *cone of confusion*. If one imagines that the two ears are both placed one a line, then all source positions on a cone that uses this line as its axis, will result in the same ITD. This is illustrated in figure 2.7.
When the cone angle θ equals 90°, the source is in the median plane, where sources from the front, above or back, all creates the same ITD. For $0 < \theta < 90$ and $90 < \theta < 180$ the ILD can be used to minimize the localization error. In the median plane, the localization is solely due to personalized spectral cues. If these spectral cues are missing, generic or unfamiliar, then front-back confusion is very likely to occur. Front-back confusion can be compensated for by visual or physical feedback [20]. If one sees the sound source in front, and at the same time

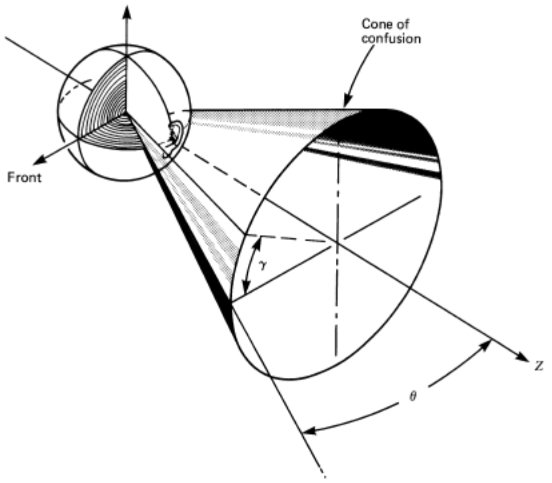


Figure 2.7: Cone of confusion illustration taken from [20]

is given a sound that can be interpreted as both front or back, our brain chooses front as the most intuitive answer. This can also be used in the opposite way as well. If the sound source is not visible, one can assume that the sound source is behind.

Sound stage distortion

The most stable localization cue is the ITD. If the ITD range used in the binaural synthesis mismatches with your own range, due to different head sizes, then the sound stage will appear wider or narrower, depending on if your head is smaller or larger than what is used in the synthesis. This will not be noticeable at small angle deviations from the median plane, but the error becomes larger as the angle moves towards 90°. For the unknown listener, the sound stage appears perfectly fine, but it will give erroneous answers in a localization task, compared to the intended source angle.

In-head localization

A very common effect with headphones is that the sound appears as it is played inside the listeners head. Out-of-head localization is possible with ordinary headphones playing a normal stereo recording, if the sound source is far right or far left, but when the sound source moves towards the median plane, this effect diminishes. By adding the acoustics effects such as reverberation and/or distinct reflections, which is commonly found in real rooms or outdoors, one can create the illusion of distance to a source, and it is easier to mentally place the source outside the head. This argument renders itself invalid, by the fact that one can experience sound coming from a direction outside of the head, when listening to a loudspeaker or other sound source in an anechoic environment. Even by using ones own HRTFs recorded in a real acoustic environment, in the binaural synthesis, one can still experience this in-head localization.

This is easily explained by the fact that our body is never perfectly still. There are micro-movements in our head, neck and torso at all times. Head movement drastically increases sound localization, because it gives our brain the opportunity to analyse the difference between two slightly different source positions, rather than a static position. When one listens with headphones and moves the head, the sound source follows the same movement, since it is physically attached to the head. This is not a natural experience for us humans, and the only sensible explanation for our brain is that the sound is in our head. There is a concept called head-tracking that can compensate for head movements. This will not be used in this thesis, since this is feature is very rarely found on commercially available headphones, and will therefore not be mentioned any further. Even if one sits perfectly still listening to a recording done with ones own HRTFs, in-head localization can still occur. This can be explained by the fact that no headphones are ideal [21], and will change the intended frequency response of the HRTFs in some way. These changes might be enough for our brain to judge the sound as deviating from what it would have received without headphones, and therefore conclude with the sound coming from inside the head.

For headphones to be optimal for binaural reproduction, they should fulfil the strict criteria for FEC (Free-air Equivalent Coupling to the ear) [21,22]. One could also reduce the effect of the headphones by adding an inverse filter to the headphone response, but even this is not bulletproof. The article *Binaural Techniques for Music Reproduction* found that the compensation filter for the headphone changed, for the same person with the same headphones, from one measurement to the next [23].

2.4 Evaluation of binaural synthesis

This section will give an short overview on different approaches when it comes to evaluating binaural synthesized material.

2.4.1 Objective or subjective? Quantitative or qualitative?

There are two main ways to test anything; objective or subjective. There are also two main types of results one can collect from such tests; quantitative or qualitative. This chapter will connect these difference to evaluation of binaural synthesis. Examples of articles will be given.

Objective and quantitative

In everyday speech, an objective statement means that the conclusions should be stated from well-known facts, and others should come to the same conclusion if using the same facts. In the world of science one has the scientific method. This method states that all results, conclusions and statements made, should be well-documented, and are only valid if the results are consistent and can be replicated by unbiased third parties. For binaural synthesis this could mean deviation in frequency response from the ideal case, where the answer could be a value in dB or a percentage value. Objective tests give quantitative data, which one can analyse statistically. The final result is usually a value, or confidence interval. An example of such work can be found in:

- *A Probabilistic Model for Binaural Sound Localization* [24]. In this article they create a computer model based on signal processing that can estimate the source location from a binaural recording. The output answer is the probability that the source is at a given location.
- *Measuring and modelling the effect of source distance in head-related transfer function* [25]. Here HRTFs are measured on human test subjects, in both near- and far-field. They are then compared to the theoretical results of a spherical head model. The results from this test consist of conclusions drawn from the trend of the numerical data, as well as a computational difference in dB.

Objective measures of binaural synthesis is not as common as subjective measures. This is because the “right” computational answer will not necessary give the best listening experience.

The benefit of an objective measure is that it is easy to implement, limit or avoid usage of the time and effort of bystanders, and one can generate large amount of data in relatively short time.

Subjective and quantitative and/or qualitative

A subjective test consists of asking the test subject about its opinion about something. One can choose if one wants the answers from a subjective test to be quantitative or qualitative. A typical quantitative answer can be obtained by giving the test subject a score range of discrete values, and ask the subject to rank something. Example: *How noticeable is the distortion in this recording in compared to the original, on a scale from one to five?.* If one wants a qualitative answer to the same question, one can let the test subject describe how the distortion was perceived. The quantitative answers can be pooled together and analysed statistically. The result of the quantitative analysis will typically reveal how noticeable the distortion is on this scale from one to five. This number can be calculated as the mean of all the answers. One might also calculate the standard deviation, and confidence intervals. These quantitative subjective answers can be compared to an objective measure of the signal at test, such as deviation from the original signal. The qualitative answers can be interpreted, and the description of the distortion can be connected to an objective measure of the distortion. The qualitative answers might also be used to create a new quantitative test that is more accurate than the first one. This time the question might be; *Rate the dominance of these effects; a) Colouration, b) Harmonics* and *c) Reduced dynamic range, on a scale from one to five.* This will give more accurate, but quantitative, answers than the first quantitative test. Some examples of subjective test for binaural synthesis:

- *Direct Comparison of the Impact of Head Tracking, Reverberation, and Individualized Head-Related Transfer Functions on the Spatial Perception of a Virtual Speech Source* [26]: The article test perceived source angle, as well of perceived realism of the spatial reproduction. The answer interface for the test subjects features two illustrations of the head, for where they have to mark where the sound is coming from. It also features a slider that goes from “Bad” to “Excellent” to rate the realism of the reproduction. The results of these tests are localization error in degrees, among with other quantitative calculated parameters.

- *Experimental Auralization of Car Audio Installations* [27]: This article models the audio installation in a car with different techniques. The test subjects were to listen to a binaural synthesis of the different techniques, and answer two “Yes/No” questions, along with a rating between 0 and 10 in how easy it is to hear the difference between the techniques. The results of this listening test were inconclusive. The only result observed was that the test subjects heard a difference between the techniques.

- *Binaural Technique: Do We Need Individual Recordings?* [28]: This article tested the localization capability of the test subject in three different conditions; real sound field, binaural with their own HRTFs and binaural with unfamiliar HRTFs. The test subject took place in the middle of a room with surrounding loudspeakers placed at different angles, distances and heights. The test subject answered by notifying which loudspeaker the sound came from, or sounded like it came from. The results were calculated as the answered loudspeaker position compared to the correct position.

Subjective tests are more common than objective for evaluation of binaural synthesis, as mentioned above. Subjective answers give a better indication of how the sound is experienced. It might also show that even though the objective measure states low quality or significant error, the subjective test can show that it is not noticeable for the listening experience. How sound is experience is subjective, and will deviate among the masses. This makes it crucial to have enough participants for a subjective test, and also ensure they are spread among different ages and genders, that are relevant to the question at test. Subjective tests are also time consuming, as one often has to compensate the participants for their time and effort, and it might be a challenge to get the right mix of participants to attend the tests.

2.4.2 Evaluation of directional hearing

Now that the general theory behind the scientific test procedures have been presented, how should or could one test directional hearing? First of all, sound is experienced differently by each individual. This is a kind of variation an automated test cannot compensate for, so testing on humans is implied. Secondly, directional hearing is no exact science, even though the mechanisms at play are known, there are large variations in the population. One can reduce the variability by using a great enough number of test subjects. Then a statistical analysis will move towards a Gaussian distribution, and one can extract the mean within a small enough confidence interval. One could also choose to collect subjective qualitative answers, but the final result will then depend on who participated in the test. It is also hard to analyse large amounts of qualitative answers in a systematic fashion. To test our ability for directional hearing, one can use one of the following test schemes:

Direct comparison between simulated direction, and correct direction

This can be done by giving the test subject a finite number of discrete directions to chose from. These directions can be represented in a number of ways. Physical loudspeakers placed around the listener can be used as answer options [28]. The positions available can be marked by numbered posters visual to the test subjects [29]. One could also use a visualization on a computer screen [26]. This visualization also features a more analogue answering, since the test subjects do not have clear discrete answer options. The use of loudspeakers to mark the positions stimulates the sense of reality of the test. Sound is commonly known to come from loudspeakers. It is therefore easier to imagine that the sound is coming from that direction, if that direction has a loudspeaker present. The numbered posters also offers something physical and visual to connect with the sound direction, but lacks the realism aspect, since a numbered poster does not commonly emit sound. The visual computer representation demands more of the test subjects' imagination to visualize the sound source direction within their own mind, and then transfer this image to the answering options on the computer screen.

The number of discrete answer options should not be too few or too many. Too few will probably give no variation between listeners, since all the other options do not match the heard source direction, leaving only one answer viable. Too many and the test subject will have a hard time deciding between the different options.

The lack of references can create errors. This especially applies when using stimuli unknown to the test subjects. Sounds like noise burst are not that common in real life, so the test subject may not have any knowledge of what a noise burst will sound like from different directions. If the test subjects are exposed directly to the test environment, and let's say the correct source direction is at 10 degrees azimuth without elevation. A test subject might answer that it comes from straight ahead, at 0 degrees, but later when the correct direction is 0 degrees, realize that he/she gave the wrong answer on the first test. These types of errors are hard to compensate. One can not give the test subjects too much insight beforehand either, as this will leave the test subjects biased. It has been shown that localization is improved if the listener has been given references of what each direction will sound like before they begin the test [29]. This was shown by first letting the unknown listener guess on their own. In the second round of the test, the subjects were to guess again, but this time they were given visual feedback of which direction was correct. When the test subjects the third round guessed, this time without feedback, the localization improved. They had then adapted to the test environment, as well as given a reference of how each direction should sound like.

Direction compared to a reference direction

A procedure known as A/B comparison with hidden reference can be used for such tests. The reference is set to be A or B at random, while the other one is the variable at test. The test subject should then decide which is which, and will often be able to rate how big the difference is, and if it is in positive or negative direction. This could be used for source localization by letting the test subject hear the reference, then let the subject listen to A and B. The listener then has to decide which of A or B has moved away from the reference position. The ranking could be by how much the source has moved. Similar errors as described in the previous section might occur if the test subject has no reference to how a given degree shift sounds like. This test can test localization in a higher resolution, than the previous test scheme. This is because it is easier for us humans to analyse the difference between two signals, than to determine the direction of a static sound.

Chapter 3

Implementation of the adaptable binaural surround sound model

This chapter will describe the theory which lies behind this adaptable binaural surround sound model. Choices made will be argued for. This model was created to be adaptable to the individual listener, that is, adaptable on a parametric level, and not as individualized as measuring individual HRTFs. Measuring individual HRTFs, as stated earlier in this thesis, is time consuming, requires special equipment and usually anechoic environment. This led to that this adaptable model modelled each of the building blocks; ITD, ILD and spectral cues, separately. These building blocks will now be described.

3.1 ITD

The ITD is, as mentioned, our most stable and consistent localization mechanism. The whole creation of the adaptable model started with the search for the optimal ITD model. Spherical head approximations are frequently used. The simplest spherical model uses a mean value for head radii, and have the ears symmetrically placed at ±90° [6]. Ears are not usually placed at ±90°, so a model where one can input an arbitrary ear placement angle was investigated [30]. Most heads are not as round as a perfect sphere, so the issue about what measure of the head should represent the radii in the model arose. Algazi, Avendano and Duda suggested a way to measure and calculate the optimal head radii [31]. It consisted of measuring width, depth and height of the head, and that the optimal radii was a result of a linear combination of these measures. Even with optimal radii, the head is still not round as a sphere. An ellipsoidal model was then investigated [1]. It also uses the measures of the head's width, depth and height, along with the displacement of the ear from symmetry of the ellipsoidal. The authors of that article showed that this model provided more accurate results compared to a real measurement, than a spherical model. The model's superiority proved best when the source was elevated. This ellipsoidal head model was chosen for the model in this thesis, because of its head-like shape and that it could be defined by a few easy-to-do measurements on a test subjects head (see figure 3.1b). The calculation of the ITD on the ellipsoidal head models is based on these five steps given in the article. Figure 3.1a illustrate these five steps.

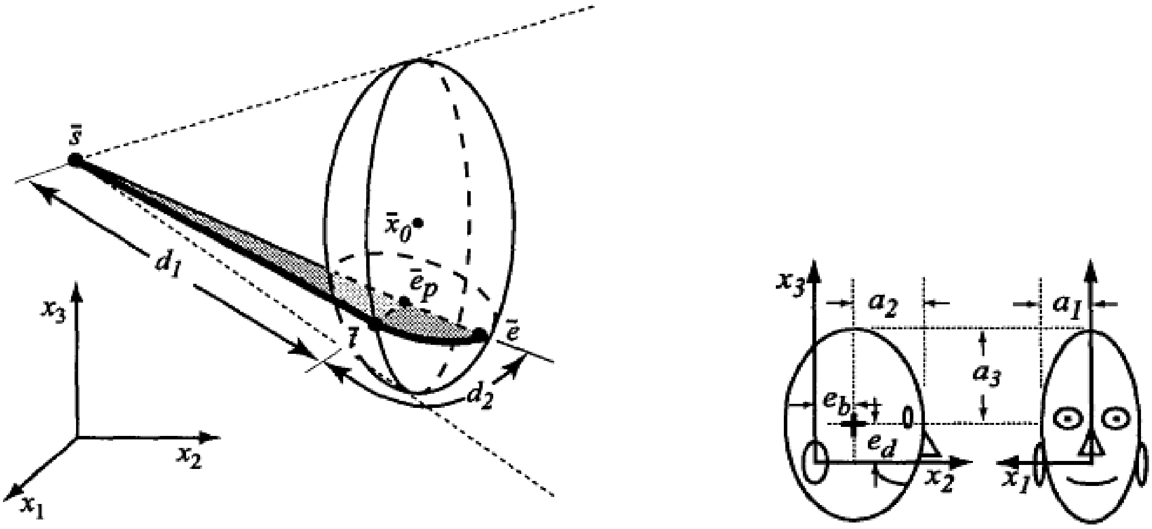
1. Find the plane P_t defined by the cone of rays from $\bar{s}$ that are tangent to the ellipsoid. The intersection of P_t with the ellipsoid defines the tangent ellipse.

2. Find the point $\bar{e}_p$ where the line from $\bar{s}$ (source) to $\bar{e}$ (shadow ear) pierces P_t.

3. Find $\bar{t}$ by assuming that it is the point on the tangent ellipse that is closest to $\bar{e}_p$.

4. Find the plane P_s passing through $\bar{s}$, $\bar{t}$ and $\bar{e}$.

5. Compute d_1 as the straight-line distance from $\bar{s}$ to $\bar{t}$, and d_2 as the arc length from $\bar{t}$ to $\bar{e}$ of the ellipse that results from intersecting the ellipsoid with P_s.



(a) Illustration of the calculation procedure to find the ITD for the ellipsoidal head model.

(b) Illustration of measurements of the head used in the ellipsoidal head model.

Figure 3.1: Illustration used in the ellipsoidal head model article [1].

3.2 ILD

ILD is, as mentioned, the dominating mechanism for localization at higher frequencies. Since heads have different sizes and shapes, there is no exact solution for this mechanism either. The approximation of a spherical head has also been used here. Lord Rayleigh derived the exact solution of the diffraction around a rigid sphere in 1904 [32]. The magnitude response as a function of frequency and incident angle is shown in figure 3.2a. This exact solution can be approximated by a much simpler model. The authors of [30] found that the following single-pole, single-zero filter created a very similar magnitude response (shown in figure 3.2b) to the response of the rigid sphere. The following shows the simple mathematics of the filter.

$$H_{HS}(\omega,\theta)=\frac{1+j\frac{\alpha\omega}{2\omega_0}}{1+j\frac{\omega}{2\omega_0}},\quad 0\leqslant\alpha(\theta)\leqslant 2 \tag{3.1}$$

ω_0 is given by

$$\omega_0=\frac{c}{a} \tag{3.2}$$

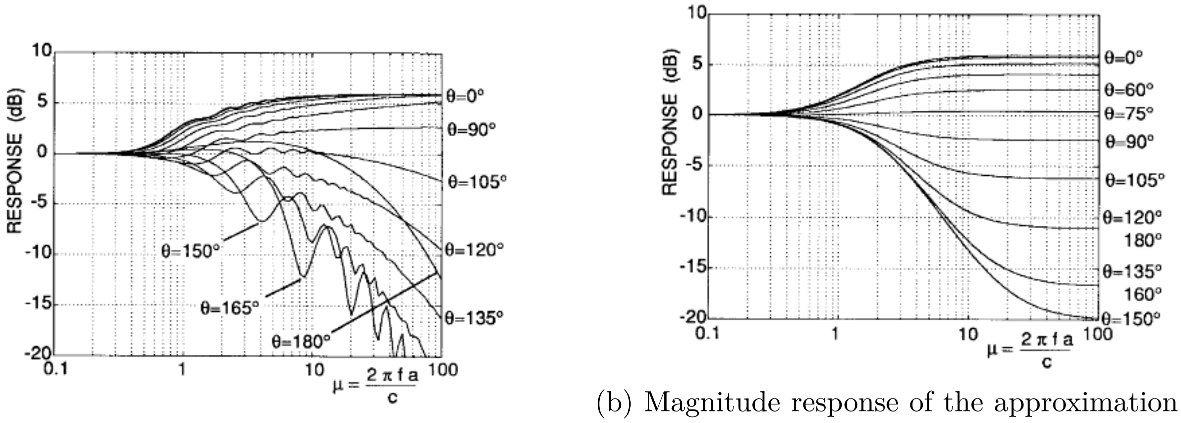
where c is the speed of sound, and a is the head radii. $\alpha(\theta)$ is given by

$$\alpha(\theta)=\left(1+\frac{\alpha_{min}}{2}\right)+\left(1-\frac{\alpha_{min}}{2}\right)\cos\left(\frac{\theta}{\theta_{min}}180^\circ\right) \tag{3.3}$$

where the authors found that $\alpha_{min}=0.1$ and $\theta_{min}=150^\circ$ produced good results. This simplified head shadow filter was chosen for the adaptable model in this thesis. A small issue was that this model needed a head radii, a, which the ellipsoidal model chosen for ITD did not have. The article [31] suggested an optimal radii based on width, depth and height measurements of the head. These are measurements that already are in the model, as they are used for the ellipsoidal. The optimal head radii was found to be calculated with the following equation.

$$a_e = w_1 X_1 + w_2 X_2 + w_3 X_3 + b \tag{3.4}$$

Where X_1, X_2 and X_3 are respectively the half-width, half-height and half-depth of the head. Similar to a_1, a_3 and a_2 on figure 3.1b. The weightings of these measurements, as well as the constant term of the equation, was found in the article to be; $w_1 = 0.51$, $w_2 = 0.019$, $w_3 = 0.18$ and $b = 32$ mm.



(a) Magnitude response of a rigid sphere.

(b) Magnitude response of the approximation of a rigid sphere.

Figure 3.2: Magnitude responses of the rigid sphere and the approximation of the rigid sphere. Illustrations taken from [30].

3.3 Spectral cues

This was the most problematic part of the model design. This is the most individual localization mechanism we have. Extensive research has been done, and is ongoing, to find well-fitting models and approximations of the pinna. There are a two main approaches to this problem; frequency band boosting, and reflections and echoes in the time domain. One of the most significant researches in the frequency boost approach was done by J. Blauert in 1970 [33]. He presented 1/3 octave band noise to the test subjects in such a way the the signal at both ears was identical. The test subjects were then asked to place the direction of the sound into three categories; front, above and back. The trend of the results was that for some frequency bands, the source was located in a given band for a significant number of listeners. These frequency bands were then named directional bands. This directional band approach has been retested on later occasions [34]. This article boosts and attenuates 6 different parts of the frequency response of generic HRTFs in search for better localization. They found that amplifying or attenuating certain bands with 12 dB gave significantly better localization with the generic HRTFs. The other approach has perhaps been a more popular subject. Solutions for this approach have been suggested in several articles [30, 35–38], with more or less accuracy. The “problem” with many of these articles, are either that their pinna-models only works for the frontal plane, or that it requires extraction of frequency peaks and notches from measured HRTFs. Both of which contradict with the intended functionality of the adaptable model in this thesis. The choice of spectral cue model fell on the directional bands of Blauert [33]. This was chosen because the article showed that this kind of frequency weighting created a sensation of front or back for significant amount of the test subjects. A plot of the sound pressure level at the eardrum from the frontal loudspeaker, subtracted the sound pressure level at the eardrum from the rear loudspeaker, is shown in this article and is shown in figure 3.3. This frequency boosting is implemented in the adaptable model in this thesis.

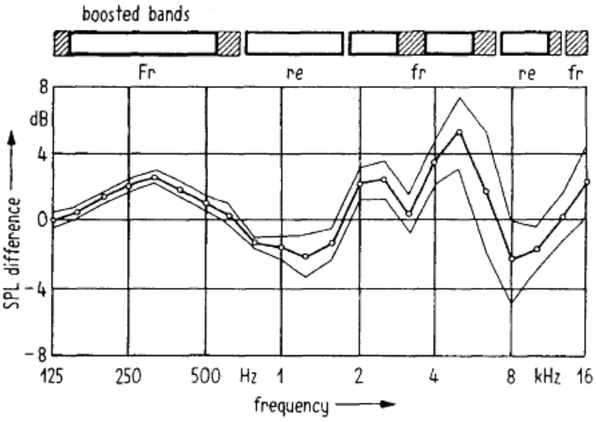


Figure 3.3: Sound pressure level (SPL) at the eardrum with loudspeaker in front minus SPL at the eardrum with loudspeaker in the rear. Directional bands shown at the top of the plot. Front (fr) and rear (re). Illustration taken from [33].

Chapter 4

Experiment

This chapter will describe how the experiment was conducted. It will describe the system that implemented the model, along with its features and graphical interface. It will describe the equipment used and the room where the tests were held, what type of stimuli was used, and why. Furthermore, it will describe who the test subjects were, what was required of them, and what information they were given.

4.1 Overview of the experiment

This experiment was conducted to see how well an adaptable binaural surround sound model would perform in localization of directional sound. To test this, a listening test was chosen. This subjective test was chosen because directional hearing, and hearing in general, is highly individual. To have something to compare the results against, the commercial success Dolby Headphones was chosen. Since headphones often create an unnatural sound stage, compared to real life sources or loudspeaker listening, the test also included a 5.0 loudspeaker setup. This had two purposes. One, to see if the test how accurate the subjects were able to localize the VBAP created virtual sources when listening over loudspeaker. Two, to see how the headphone models performed compared to the loudspeaker setup.

4.1.1 Test specifications

To test the subjects' ability to localize sound, and to be able to convert that localization into a quantifiable answer, a uniform spatial resolution of 10 degrees was chosen. This resolution is fine enough to not cause the correct direction to be to obvious, but coarse enough so that the test subjects would be able to decide with confidence. The positions were limited to the horizontal plane only. Loudspeakers were placed to represent the discrete positions. Due to a limited number of loudspeakers, the test was made one-sided. Symmetrical sound localization assumed. By random, the right side was chosen. 19 numbered loudspeakers placed from 0 to 180 degrees, and numbered 1 to 19, was placed on a half circle around, and on the right side, of the test subject. The half circle had a radii of 1.93 m, from the center to the face of the loudspeakers. The loudspeakers were placed on stands, so that the height from the floor to the center of the loudspeakers were 1.07 m. This placed the loudspeakers in the head height for a seated position.

4.1.2 Test directions

Ten directions were chosen among the 19 as source directions. This limitation created the opportunity to represent each of the ten source directions multiple times in each test, without the test becoming too time consuming. This is advantageous because one can see if the same test subject will answer the same direction, or not, when it is repeated. It will give a measure on how easy it is to determine a source direction, and the consistency of the test subjects, even if the test subject answered the same, but wrong, for all repetitions. This would reveal that for this test subject, the source sounded like it came from a different direction than intended, and did so consistently. This is an effect that one could possible compensate for in a model, if the effect should occur. Three repetitions of each direction was chosen, resulting in a total of 30 directions per test. With the three tests; loudspeakers, Dolby Headphones and thesis model, made the total of 90 directions to be tested per test subject. A listening test took about 20 minutes, including the measuring of the subject's head, and explanation of the test procedure. The ten selected source directions are shown in table 4.1, with both numbering and azimuth angle.

Table 4.1: Source directions for the listening test.

No.	1	3	4	7	8	10	12	16	17	19
Angle	0	20	30	60	70	90	110	150	160	180

These ten directions were chosen for several reasons. They are spread out over the half circle. Some directions come in front-back pairs, such as 1-19, 3-17, 4-16, and 8-12. 90 degrees, no. 10, is also included, since this theoretically is a weak spot for VBAP. Lastly no. 7, which does not make up any pair, and lies in the panning region between front right loudspeaker, and right surround loudspeaker. Figure 4.1 shows a picture of the actual loudspeaker setup, and listening position, used for the listening tests.

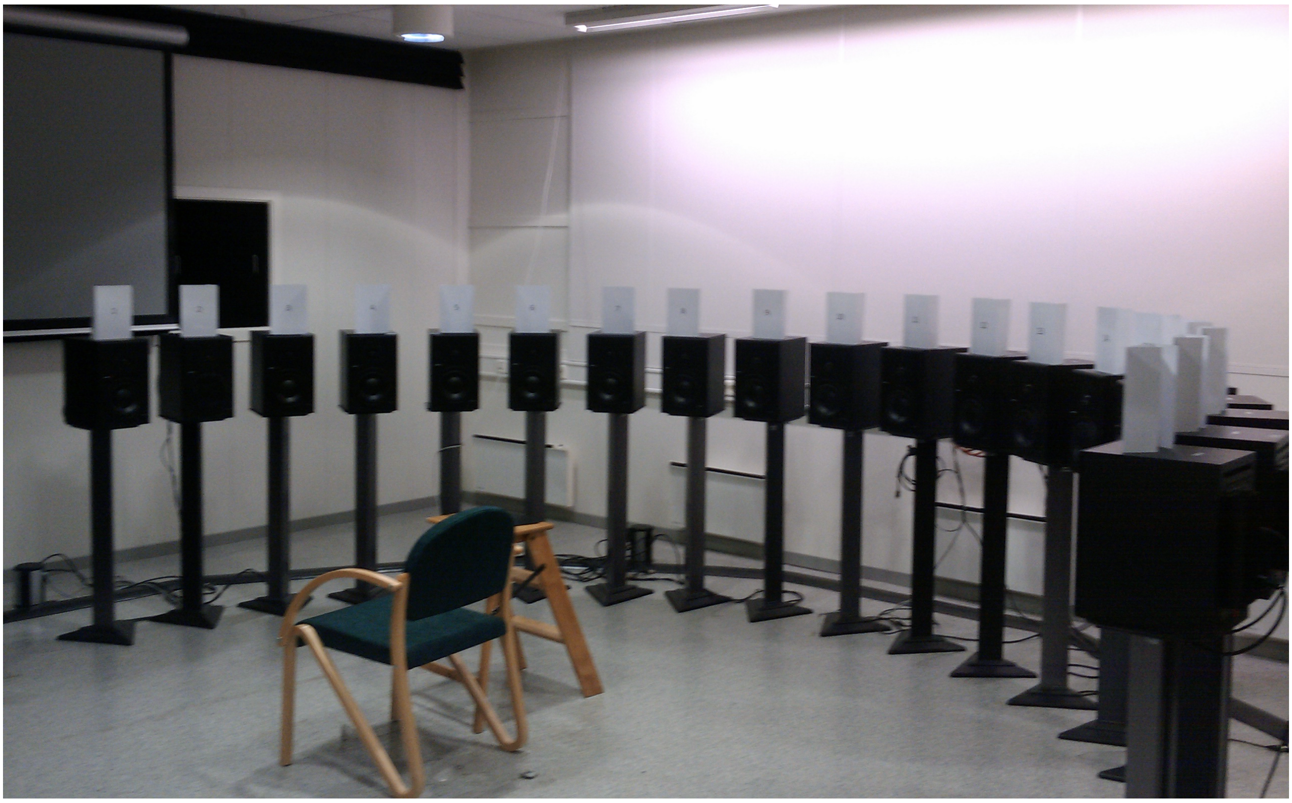


Figure 4.1: Picture of the loudspeaker setup in the listening room. The white squares are paper sheets with numbers ranging from 1-19.

4.2 Experiment testing software

This section will give an overview on how the aforementioned chosen theory and models were implemented. The chosen programming language for this implementation is MATLAB. This section will mention names of different MATLAB functions. Appendix A contains descriptions for each function, what it takes as input, and what is given back as output. For a complete view of the functionality, one can view the complete files used on the attached CD. The different theoretical parts and models are implemented separately in different blocks, or MATLAB functions if you like. These blocks are then combined together in what in thesis is called the feedback system. This is a complete system with all the needed functionality and a graphical user interface for the test subjects. The coarse structure and functionality of this system is illustrated in figure 4.2. A short description on the functionality of each block inside the grey frame follows. The rest will be explained in later sections of this chapter.

4.2.1 Listener

This class offers three main features;

- General data such as test subject anonymous ID, age and gender.
- Measured data such as head width, depth, height and ear placements.
- Random generated azimuth values for each test, with a matching storage for the answers.

This class structure was chosen to store all the needed data for the analysis of the results. Each test subject resulted in a *Listener* class object to be placed in a vector. In this way, a script could be made afterwards to iterate through the data, to easily extract and compare values. This was found to be highly effective and convenient way to store the data, and made the analysis easy. This also kept track of the relationship between what the correct azimuth was, and which azimuth was answered. This offered a complete randomization for each test subject, without any pattern or clues. For further details about the Listener-class, see appendix A.2.

4.2.2 VBAP for 5.0

This block implements the vector base amplitude panning described in section 2.2.1. This block takes the azimuth for the virtual source to be generated as input. It then finds out where this source is placed in comparison to the 5.1 loudspeaker setup illustrated in figure 2.5, and calculates the gain of the two nearby loudspeakers. The “.1” is omitted, since it will not contribute to the directional playback. More information about this block can be found in the description of the *intensityPan* function in appendix A.1.

4.2.3 ITD Ellipsoidal model

This block uses the information about head dimension and ear placement stored for the current test subject, along with the azimuth of the virtual source, to calculate the ITD. The ITD is calculated using the ellipsoidal head model described in section 3.1. This calculation is made with the use of the MATLAB functions; *ellsection*, *elltan*, *ellipsearc* and *mineldist* described in appendix A.1.

4.2.4 Head shadow filtering

This block calculates the optimal head radii based on the equation 3.4 with weighting given below the equation. This radii is used to implement the one-zero one-pole head filter for the given source azimuth angle, as described in section 3.2. This block is implemented in the function *headShadow* described in appendix A.1.

4.2.5 Spectral cues

This block checks if the source is in front of the listener, or behind the listener. It then adds frequency characteristics presented in figure 3.3 if the source is in front of the listener, or the inverse characteristics if the source is behind the listener. The added frequency characteristics are scaled based on where the source is: scaling value one if the source is right in front of or right behind the listener, and value zero if it is at ear angle. This is done because the ITD and ILD provide enough clues at the side, but become weaker when the source moves towards the median plane. This block is implemented in the function *directionalBoost* described in appendix A.1.

4.2.6 Generate binaural

This grouping of blocks made the coding more convenient. This grouping is implemented in the function *generateBinaural* described in appendix A.1. It uses the function described in the blocks; ITD, head shadow and spectral cues.

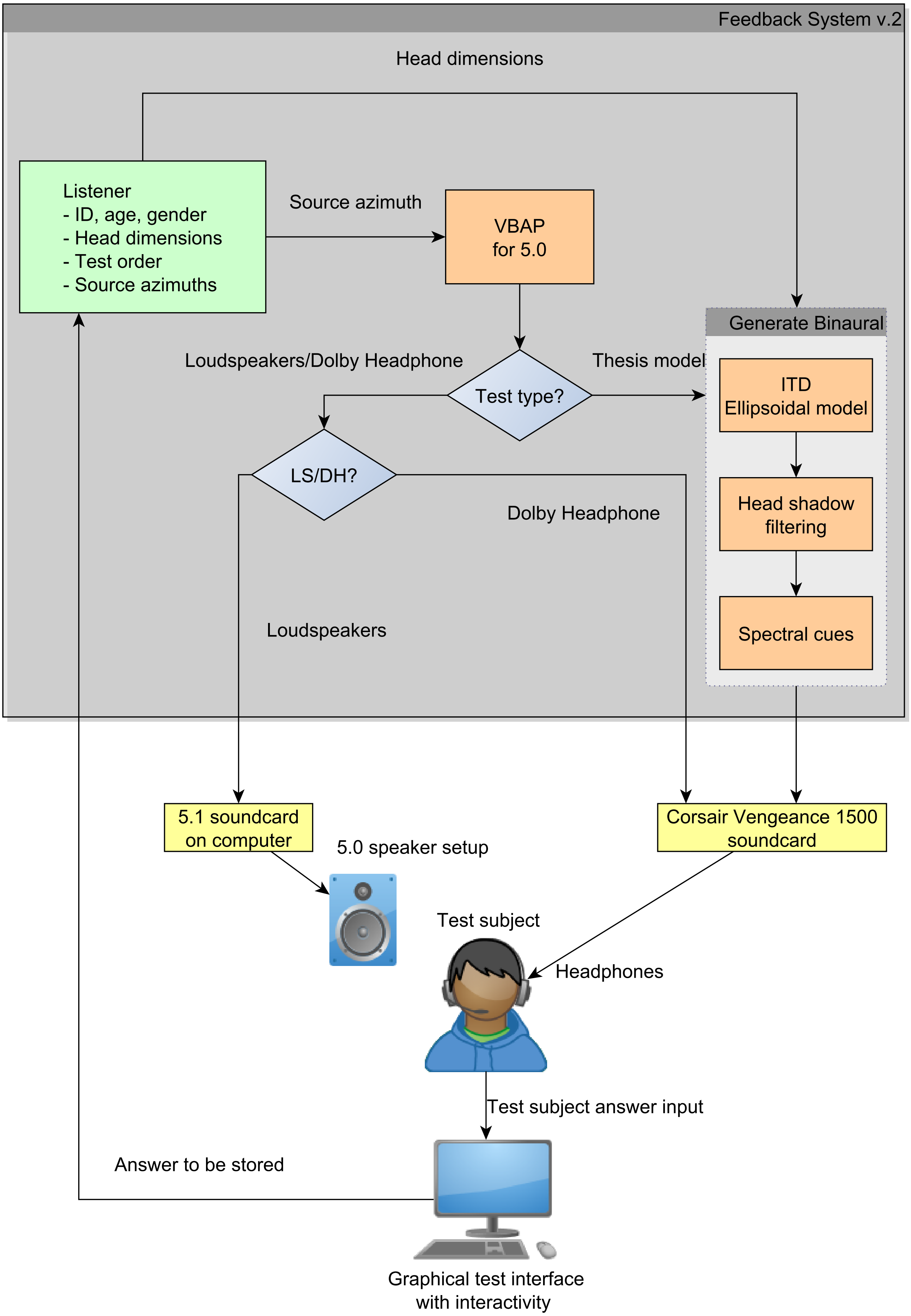


Figure 4.2: Flowchart of the basic functionality of the feedback system. Green represents a class structure, orange represents functions and yellow represents hardware.

4.3 Hardware and listening room

This section will present the equipment and listening environment used during the experiment.

4.3.1 Loudspeakers and sound card

The loudspeakers used were of the type dynaudio acoustics® professional monitoring system BM6A. These have the opportunity for some frequency response modifications by settings on the loudspeakers itself. These settings were set to the following; HF trim dB at 0 dB, LF trim dB at 0 dB and level dBm at +4. These loudspeakers are active, and do not need any pre-amplification of the signal. The signal was supplied by the on-board soundcard. The motherboard was of the type ASUS P8 H61 Rev. 3.0. The on-board soundcard used ALC887 8-channel High Definition Audio CODEC.

4.3.2 Headphones

The headphones chosen were a pair of Corsair® Vengeance® 1500 7.1 USB Gaming Headset. This headset is mid-range when it come to price at the time this thesis was written. It comes with a USB soundcard with implemented Dolby DSP technology, such as Dolby Headphones among other features. Figure 4.3 shows an image of the headphones.

Figure 4.3: Corsair® Vengeance® Dolby 7.1 USB Gaming Headset. Image from [39].

These headphones were chosen for a number of reasons. First of all, they offered the Dolby Headphone technology, which was necessary for use in this experiment. Secondly, they have received great reviews and come with good specifications [39]. Lastly, they are in an affordable consumer price range valid for most people looking for a decent, but not too expensive, headset. The USB sound card comes with a driver software which allows the user to change between different output modes; bypass, Dolby Headphones and 7.1 Virtual Speaker Shifter. Bypass-mode does not add any kind of processing. In this mode, the headset works as a normal stereo headset. This mode was used for playback of the adaptable binaural model in this thesis. 7.1 Virtual Speaker Shifter offers the opportunity to move the eight loudspeakers around the listener, both in angle and distance. This mode was not used in this experiment. The "Environment Size" *Studioa* was chosen in the Dolby Headphone-mode. This was chosen to best represent the room used for the listening tests. The software also give the user the opportunity to choose from preprogrammed equalization setups, along with the opportunity to create their own equalization setups. The equalization was set to 0 dB (flat) for this experiment. The headset's performance, such as frequency response or fulfilment of the FEC (Free-air Equivalent Coupling to the ear) criteria [21, 22], was not tested. This was not tested or adjusted because the experiment is based on hardware, software and content, that are, or easily could be (such as the thesis model), available for the common consumer. Individualized inverse filtering of the headset response is not such a feature.

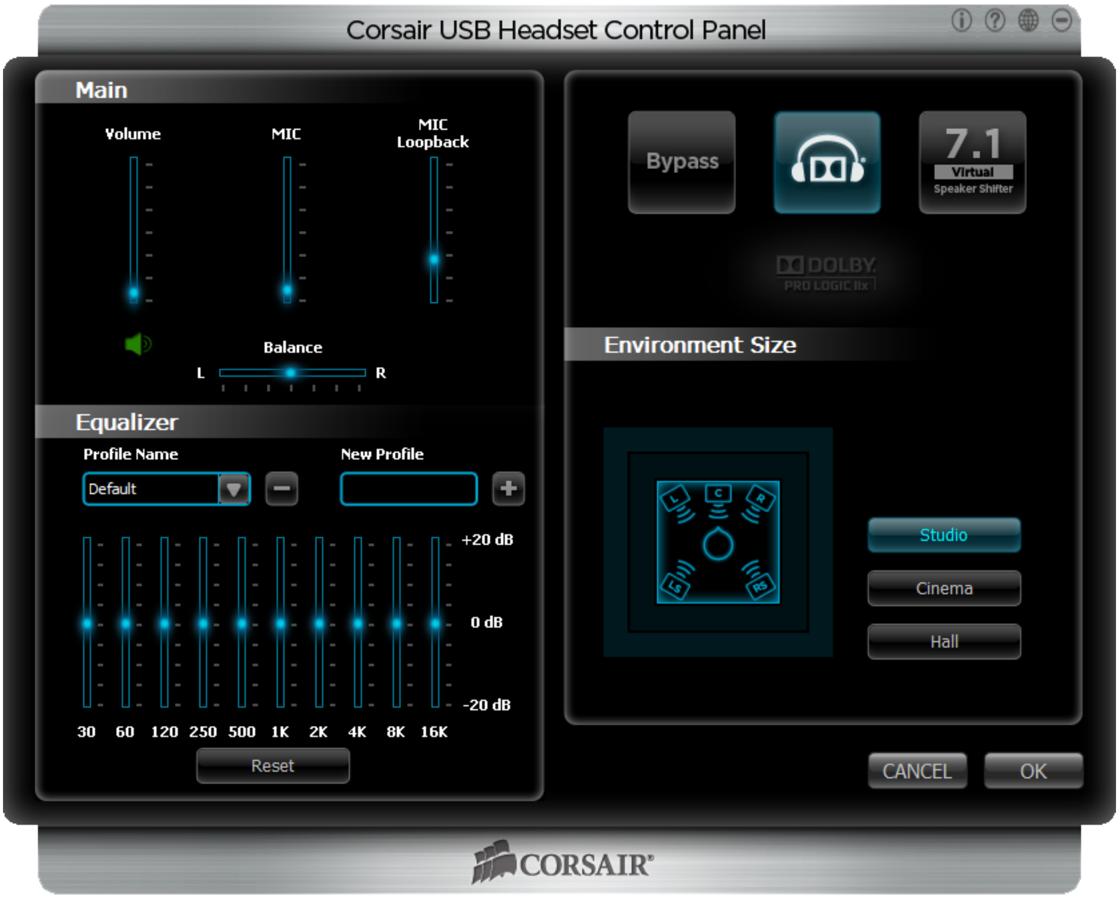


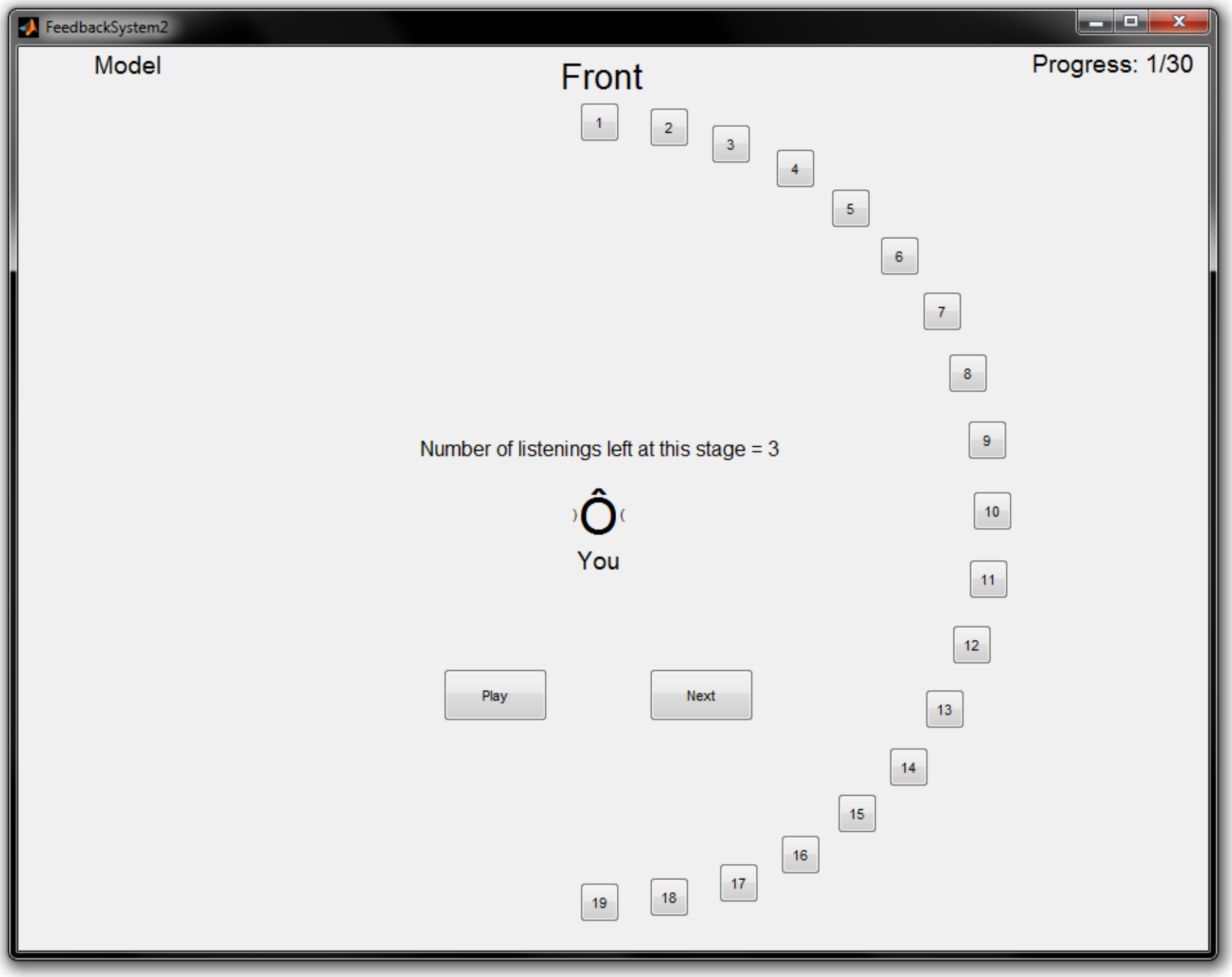
Figure 4.4: Screenshot of the Corsair USB Control Panel. Currently showing the Dolby Headphone mode, along with the equalization option.

4.3.3 Listening room

The room used for the listening tests measures 7.3 x 5.85 x 2.75 metres. This room is used for evaluation of surround sound and audiovisual material daily. It offers a low and flat reverberation time, and high diffusivity.

4.4 Graphical test interface and interactivity

The feedback system was created in GUIDE (MATLAB's Graphical User Interface Design Environment). Figure 4.5 shows a screenshot of this interface. The interface represents the actual loudspeaker placement and listener seating of the test area. Each numbered box represents an actual loudspeaker. The test subject was given the opportunity to play the test stimuli three times for each direction, at each test. Feedback from the test subjects showed that this feature was found to be very helpful. The answer is given by clicking the numbered box representing the loudspeaker the sound came from, then hit *Next* to continue to the next source direction. All actions could be activated with a simple computer mouse, which was placed on a small table to the right of the test subject. When the *Progress*-counter reached 30 for one test, the test operator was reminded to change the sound output settings on the computer. When the test is done, a *Close* button appears. Clicking on this button will save the answers to the current *Listener*-class object. This interface was shown on the grey projector screen visible to the upper left on the picture in figure 4.1.


Model

Progress: 1/30

Number of listenings left at this stage = 3

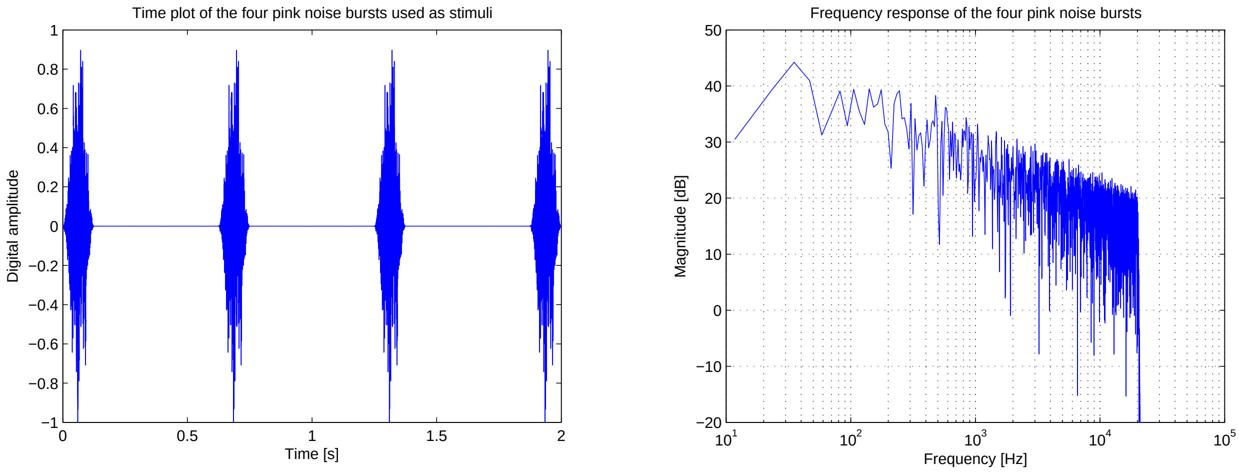
Figure 4.5: Listening test answering interface, created in MATLAB GUIDE.

4.5 Stimuli and sound volume

The choice was simple when it came to selecting test stimuli; pink noise. Pink noise has a -3 dB decay, which provides equal amount of energy in all octave bands. It also contains all frequencies, which is essential for testing all the hearing mechanisms. Pink noise is also easier to listen to, compared to white noise, due to the -3 dB decay. A former master's thesis [29] did a pilot test on stimuli for listening tests. The most preferred stimuli was bursts of pink noise. Figure 4.6a shows the stimuli used for this thesis. It has the same burst length, and pause length in between bursts, as the preferred stimuli in the mentioned thesis. 125 ms of burst, separated by 500 ms of silence. The only difference, is that these bursts have been filtered with a hamming-window[1], making the onset and offset smoother. This was done to achieve a more comfortable sound, and also to reduce the potential for distortion in the loudspeakers and headphones. Figure 4.6b shows the frequency content of these four bursts. One can clearly see the -3 db decay. The pink noise is also passed through a band pass filter from 20 Hz to 20 kHz. This was also done to reduce the potential for distortion in the loudspeakers and headphones.

One might argue that pink noise bursts are not common sounds in every day life, so the test subjects had no familiarity with it, and would need time to adjust to the sound. The positive part is that pink noise is such a neutral sound, and most test subjects will be completely unbiased for this sound, compared to other sounds that they could have a relationship with.

The sound volume the sounds were played at were only registered as output settings on the loudspeaker or headphone sound cards. For loudspeakers, the output volume was set to 65 out of 100. For headphones 20 out of 100. A few subjects requested a slightly higher volume for the headphone tests, and found the setting 25 out of 100 adequate. Comfort was prioritized over having the exact same sound pressure level for loudspeakers and headphones.



(a) Time plot of the four pink noise bursts used as stimuli.

(b) Frequency response of the four pink noise bursts.

Figure 4.6: Time and frequency representation of the listening test stimuli.

4.6 Test subjects

20 test subjects participated in the experiment; 15 male, and 5 female. All test subjects were university students in the age range from 23 to 31 years. Complete age distribution can be viewed in figure B.1. Their hearing ability was not measured in any way before participating in the listening test, but they were asked if they had any problems with hearing in everyday life, which no one felt that they had. They were not tested, because an average consumer of a potential binaural surround sound product, would not be required to pass any hearing test before buying the product.

[1] A hamming window is a time window with a defined continuous shape, starting and ending with the value zero. This reduces the effects of the onset and offset of a signal, which can create unwanted frequency effects.

4.6.1 Head measurements

Before the test subjects commenced with the listening test, some physical measurements were made. Head width, depth and height were measured, along with ear placement back and down, and the height of their ear canal entrance while seated. Values for each of the test subjects can be found in figures B.2 (head measurements), B.3 (ear placement) and B.4 (seated height). The head size and ear placement measurements were made with the tool pictured in figure 4.7. This tool was made for these tasks in mind, and has an accuracy of ± 2 mm. There are other ways of measuring these sizes, such as digital photography or encasement of the head [31]. The pictured device was chosen for its quick and convenient execution of the measurements. It also gave an acceptable degree of accuracy. Height of the ear canal entrance while seated was measured with an ordinary hardware-store meter ruler.

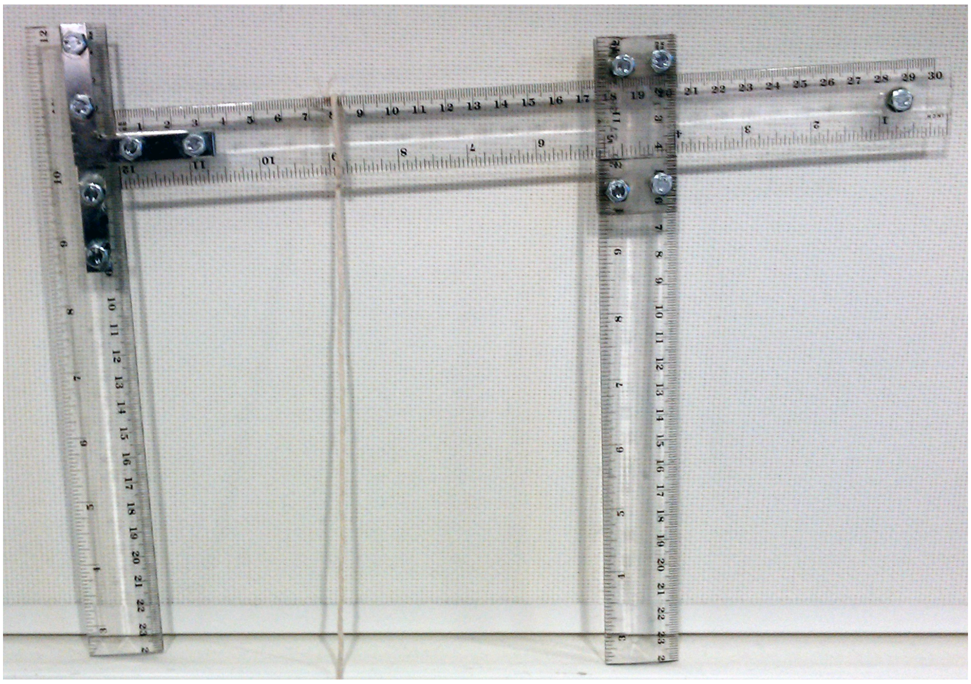


Figure 4.7: Tool used for measuring head size and ear placement.

4.6.2 Information

After the different measurements on the test subjects, they were given the following information and terms about the test.

- There are three test; one with loudspeakers and two with headphones. The order in which you are going to take these tests will be decided at random.
- Each test consists of 30 directions. The directions can correspond to any of the loudspeakers, and one direction can appear more than once. This is also generated at random.
- You have the opportunity to listen to each direction three times. This is done by pressing the *Play* button.
- When you have decided on your answer, click on the button corresponding to the loudspeaker the sound came from. Then press *Next* to continue. When an answer is given, you can not change it, and you must continue.

- During the loudspeaker test, please keep facing forwards when playing the sound. You might turn your head afterwards to look which loudspeaker direction the sound came from.
- Speak up if the sound volume is to high or low, and it will be changed to a comfortable level.
- The data and results corresponding to you will be anonymized, and you might choose to terminate this test at any time.
- You will be rewarded one gift certificate for a cinema viewing, should you choose to complete this listening test.
- Any questions? If not, let's start.

4.7 Randomness

The test order and the ten different source directions, repeated three times, for each test, were all randomized. This was done by filling the arrays with the total amount of values it should contain, and then shuffling these around. This was done by the *Listener*-class function *shufflePos* found in A.2. The built-in function *randi* in MATLAB was used to generate random matrix indexes to be swapped. After all the experiments were concluded, it was discovered that this function creates the same pseudo-random numbers for each time MATLAB is started. This was not known beforehand. The 20 listening test were spread out on different days, with one to a few test subjects participating each time. This means that there is a non-uniform distribution of test order and stimuli. In other words, a larger part of the test subjects received the same test order and stimuli. This is not optimal, but since the test subjects were told that it was all random, they might not have seen any reason to discuss their results with anyone else. Hopefully, it will not affect the final results in a significant way. This randomizing error should be corrected in future experiments.

Chapter 5

Results

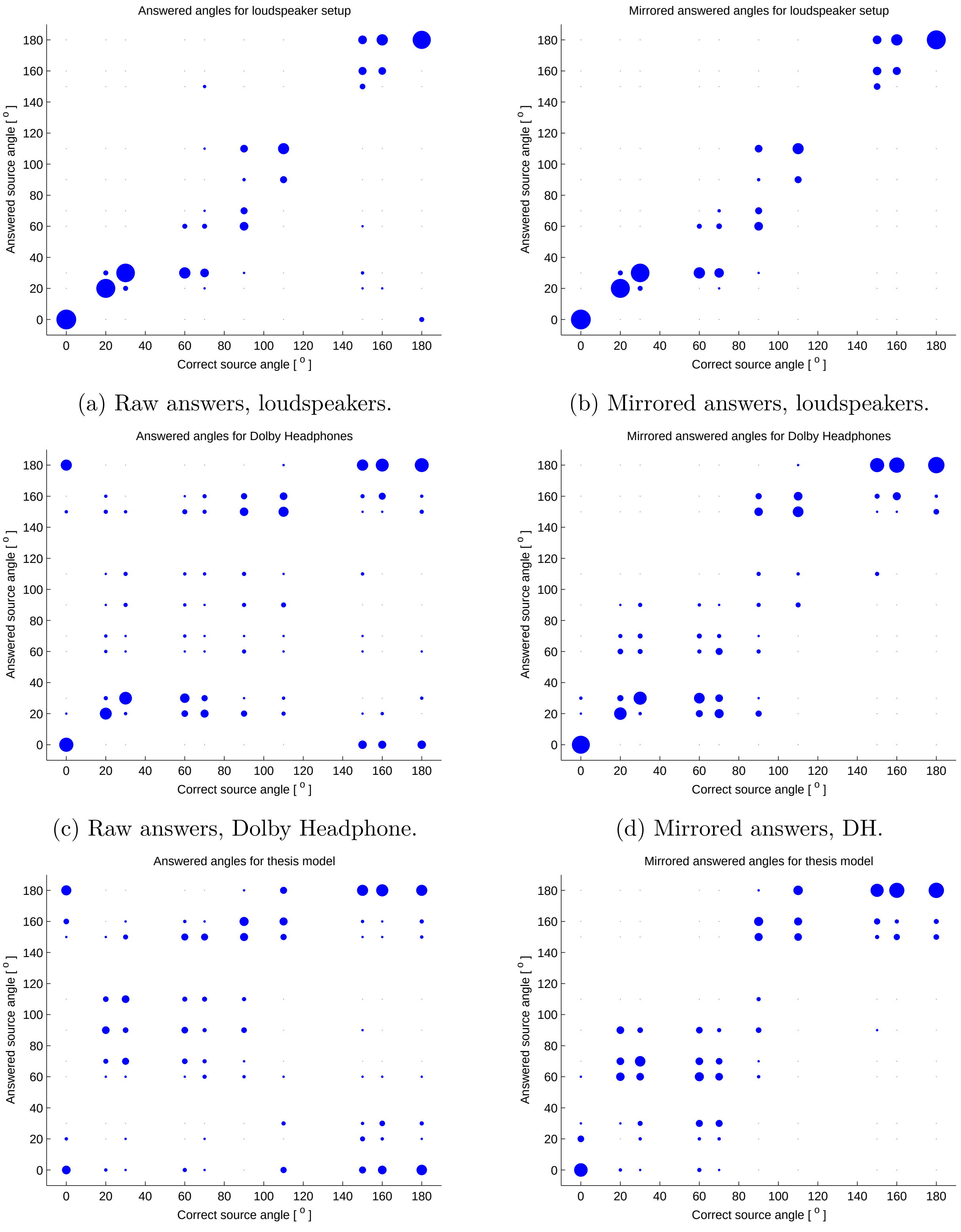
This chapter will present the results from the listening tests. The results will be presented as different plots showing raw or calculated data. These graphs or calculations are based on the following amount of data points. 20 test subjects participated in the listening test. The listening test consisted of three different test types; loudspeaker setup, Dolby Headphone and thesis model. Each test type had 10 unique source directions, with 3 repetitions for each direction. This sums up to 600 data points per test type. 60 data points per unique source direction, per test type. 90 data points per test subject.

5.1 Raw and front-back corrected results

Figure 5.1 visualizes the raw and corrected results from the listening tests. These plots are called bubble plots. Each answer combination is covered and unique. If a combination has been answered, the occurrence of this combination is increased. The larger the bubble, the more occurrences at that combination. The raw answers for the three test types; loudspeaker setup, Dolby Headphone and thesis model are given in respectively figures 5.1a, 5.1c and 5.1e. The front-back confusion corrected answers are corrected by mirroring them around 90 degrees, if there have been a front-back confusion. For example if the correct angle was 30 degrees, and the test subject answered 130 degrees. Then it is corrected to its mirrored angle, which is 40. The angular error is now only 10 degrees, and not 100 degrees. This is done by the function *frontBackConfusion*, which is further explained in appendix A.1. This front-back confusion correction is based on the strict assumption that a visual feedback in front of the test subject, would rule out all front-back errors. The visual feedback could be the movie the sound belonged to, or the game play that created the sound. These front-back confusion corrected answers will be addressed as “mirrored” from here on. The mirrored answers for the three test are found in respectively figures 5.1b, 5.1d and 5.1f. Table 5.1 gives the number and percentage of how many of the test answers were mirrored for each test type.

Table 5.1: Occurrences of mirrored answers for the three test types. Value as number of occurrences, and occurrence in percentage. Values calculated from a total of 600 per test type.

Test type	**No. mirrored**	**% of mirrored**
Loudspeaker setup	28	4.7
Dolby Headphone	207	34.5
Thesis Model	325	54.2



(a) Raw answers, loudspeakers.

(b) Mirrored answers, loudspeakers.

(c) Raw answers, Dolby Headphone.

(d) Mirrored answers, DH.

(e) Raw answers, thesis model.

(f) Mirrored answers, thesis model.

Figure 5.1: Bubble plot of raw answers (left column) and mirrored answers (right column) for loudspeaker setup(top), Dolby Headphone(middle) and thesis model(bottom). Answered source angle, as function of correct source angle. Size of bubble represent how many occurrences that source direction combination have received.

5.2 Mean and standard deviation

Figure 5.2 shows the mean of the mirrored answers for each direction, for each test. The shaded area around each curve corresponds to the 95% confidence interval. Figure 5.3 shows the mean of the mirrored answer minus the correct answer, for each of the tests. Table 5.2 gives numerical values for the total mean values, and upper and lower 95% confidence limits, for the difference between mirrored and correct, for each test. Figure 5.4 shows the standard deviation for the mirrored minus the correct answer, for each angle and each test. Table 5.3 gives numerical values for the total standard deviation for each test type.

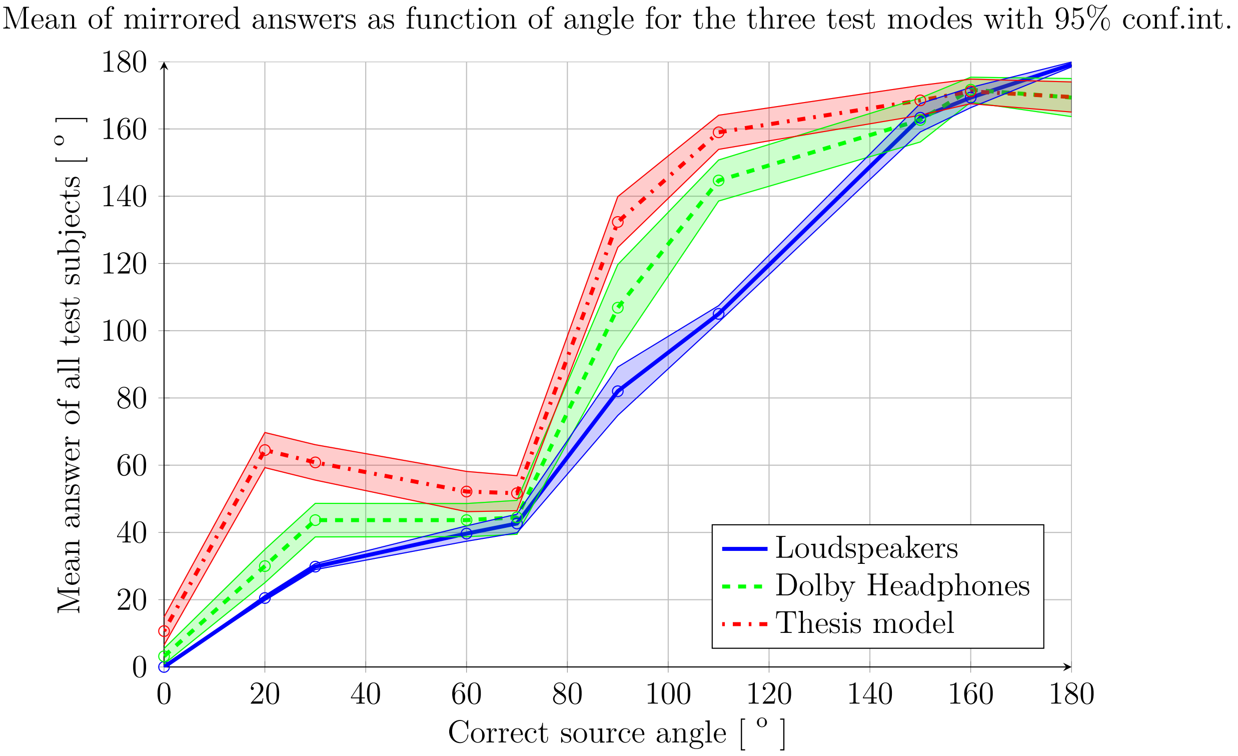


Figure 5.2: Mean of mirrored answers, with 95% confidence interval for the three test types.

Table 5.2: Total value of mean and 95% confidence interval limits for correct answer minus the mirrored answer, for the three test types. Values are in angle degrees.

Test type	**Lower**	**Mean**	**Upper**
Loudspeaker setup	-5.2	-3.9	-2.5
Dolby Headphone	2.7	5.0	7.4
Thesis model	14.6	17.0	19.5

Table 5.3: Total standard deviation for the three test types. Values are in angle degrees.

Test type	**Standard deviation**
Loudspeaker setup	16.8
Dolby Headphone	29.5
Thesis model	30.5

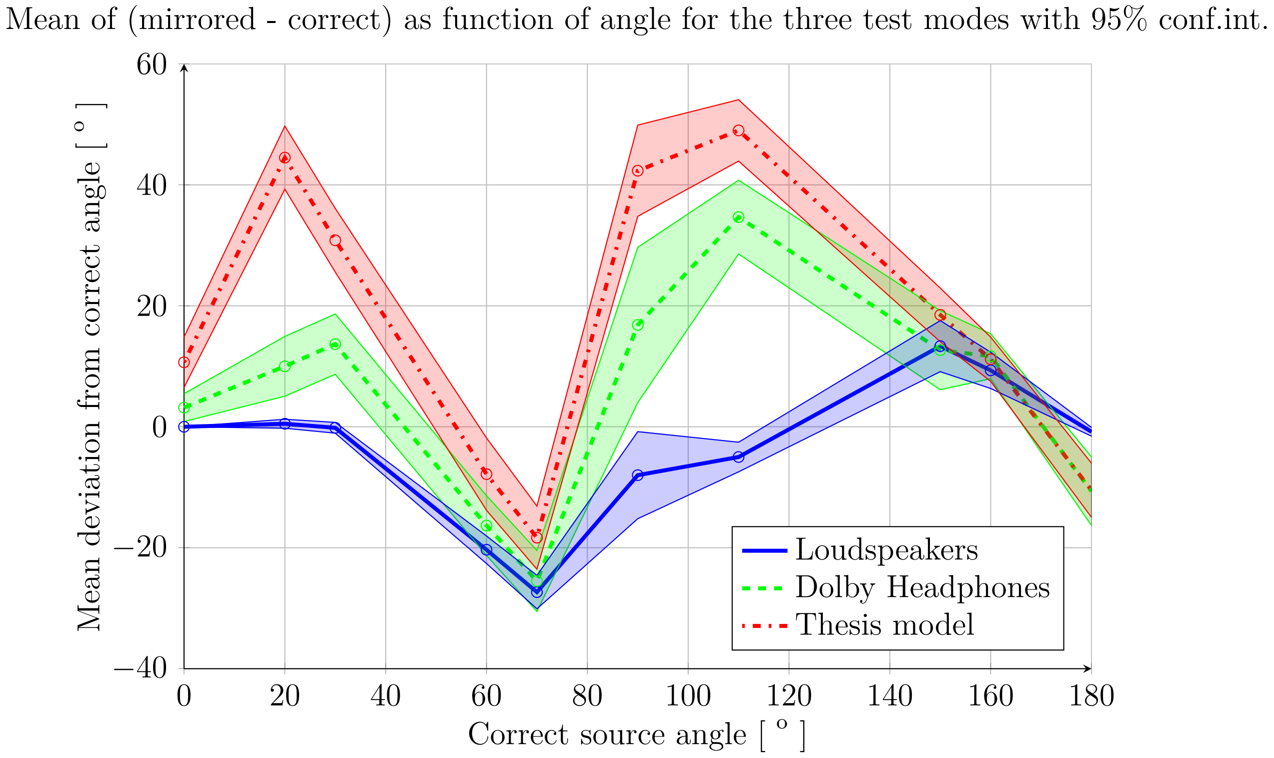


Figure 5.3: Mean of mirrored answer minus correct answer, with 95% confidence interval for the three test types.

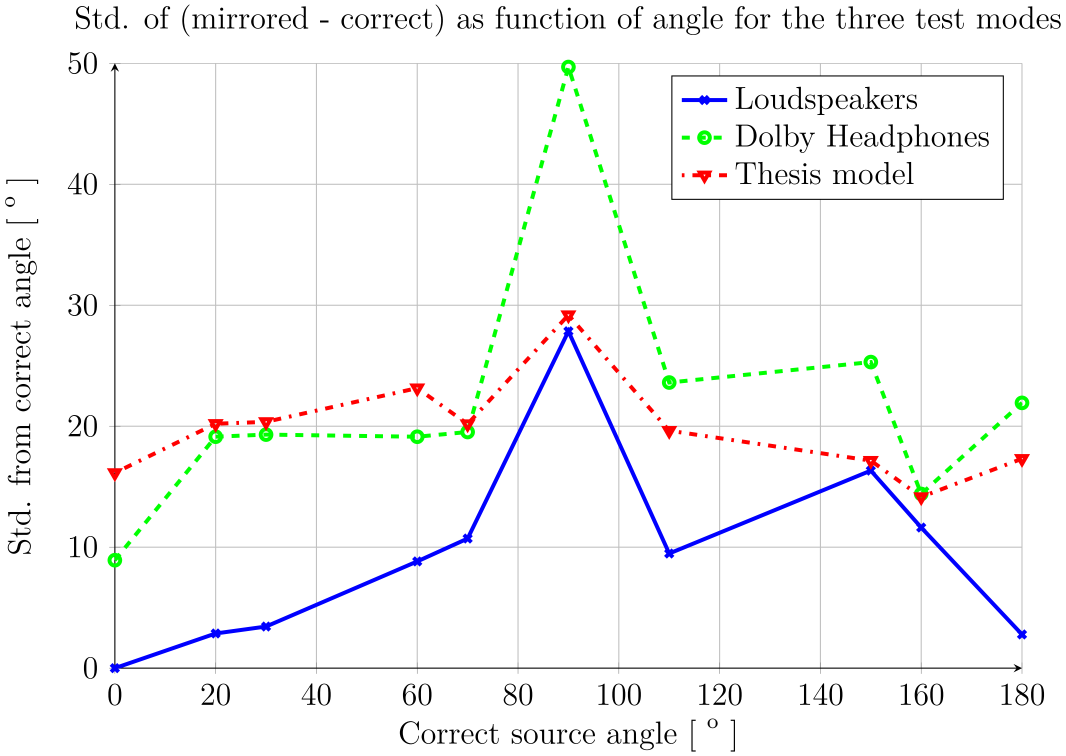


Figure 5.4: Standard deviation of mirrored answers for the three test types.

5.3 Test subject consistency

Each unique direction got repeated three times for each test, for each person. This makes it possible to calculate the standard deviation for each test subject. This is done by calculating the difference between the answered source direction, and the mean of the answers given for the three repetitions for that source direction. This enables us to see how much these differences vary in total for each test subject. If a test subject gets zero standard deviation, it means that the test subject answered the same source direction all three times for all directions. This result could be achieved, even by answering horribly wrong, as long as the test subject answered the same wrong direction for all repetitions. In other words, this value describes the consistency of the test subjects. This could also be interpreted as how hard or easy it was for the test subject to get a clear sense of direction, even if it was the wrong direction. A low standard deviation, but wrong answer, could mean that the model at test produce a consistent error, which can be compensated for in the model design.

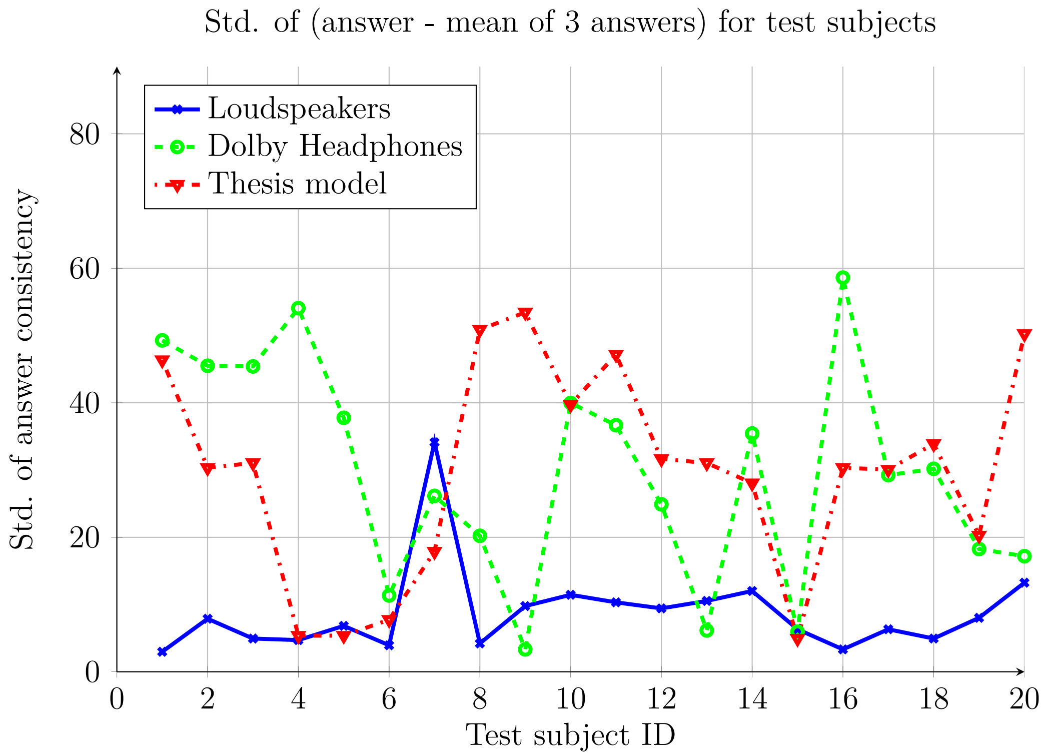


Figure 5.5: Standard deviation for each test subject. Mirrored answer minus the mean of the three answers for that source direction, for each test type.

Chapter 6

Discussion

This chapter will discuss the different results obtained from the listening tests. What does the results tell us, and why are the results as they are? These questions, and more, will be answered in this chapter. Some figures from the result chapter will be reposted here, for convenient viewing while reading the discussion.

6.1 The raw results

When analysing the results, one must keep in mind what would be the perfect results. For the plots in the figures following each subsection (originally figure 5.1), this would be a straight line with only big bubbles starting at 0,0 and ending at 180,180.

Loudspeakers

From the raw results for the loudspeaker setup in figure 6.1 one can see that the results are fairly collected around the line that would be the perfect result. All the test subjects have gotten the source direction 0 correct for all repetitions. There is also very little error for 20 and 30 degrees. Also at the end, the results are fairly good, especially at 180 degrees. One can see that the directions close to 180 degrees have often been localized as 180. This could be because the test subjects were presented with one or more 150 and 160 degrees stimuli before a 180 degree stimuli, and therefore thought that it was as far back as it could be, and then realize their error when they heard a 180 stimuli. It might also be due to lower spatial resolution for sounds from behind, or that the intensity panning has a weakness in this area. The middle section however, has much more spread. Answers such as 30 or 110 degrees occur more than others. These directions have the actual playing loudspeakers used for the intensity panning. This area between these two loudspeakers is a known weakness for the 5.1 loudspeaker setup. This is due to the cone of confusion mentioned in section 2.3.5. The cone of confusion makes as if the ± 110 degree loudspeakers, actually were placed at ± 70 degrees. This is based on a low frequency approximation, where the ITD is the leading localization mechanism. This limits the possibilities of creating virtual sources near 90 degrees. It is not possible to create a virtual source at 90 degrees from either the front or the back, without actually placing a real source there [40]. All in all, the raw results from the loudspeaker setup is fairly good. This is also expected, since the test subjects were able to use their own ear mechanisms without alterations for this test.

Dolby Headphone

The raw results for the Dolby Headphone in figure 6.2 have much more spread for all directions, compared to the loudspeaker setup. The direction of 0 degrees has about a 60/40 spread among

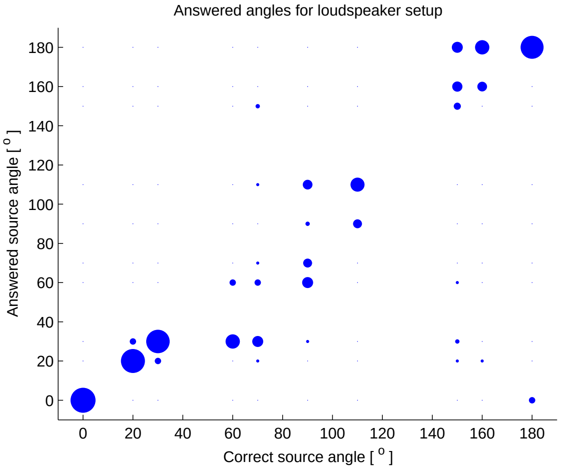


Figure 6.1: Raw answers, loudspeakers.

0 and 180 degrees. The 20 and 30 degree direction is localized more correct, than wrong, but still have single answers spread out among almost every direction. The three directions at the back also have about a 60/40 spread with their mirrored direction in front. The middle section is very widely spread, and is in general localized either closer to the front or the back, than what they should be. The direction answered more than others for this this section, does not directly correspond to the thought loudspeaker positions of the virtual 5.1 system. 110 degrees has nearly no answers at all, while 150 has many. In the front, the answers are about evenly spread among 20 and 30 degrees. Almost every source combination, except 0 and 180 degrees, have been answered for the middle section. This might indicate individual difference with the model used in Dolby Headphone, as well as problems in the model in general. Dolby Headphone has virtual room reflections and reverberation. These might also be the reason for this almost uniform spread, making the sound more diffuse than direct.

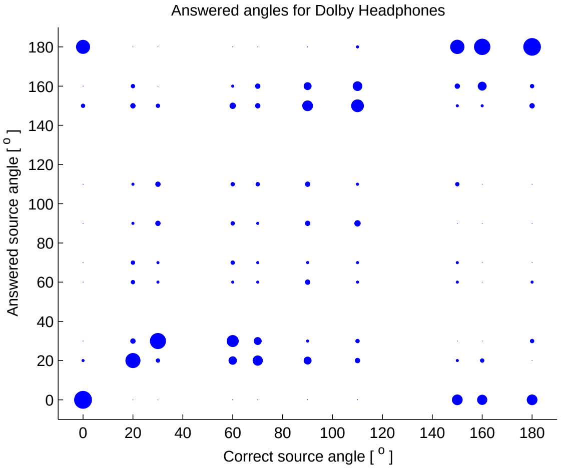


Figure 6.2: Raw answers, Dolby Headphone.

Thesis model

The raw results for the thesis model in figure 6.3 have an even worse spread than Dolby Headphone. The trend leans towards the back of the half circle, with only one correct answer for 20 and 30 degrees. The front-back problems for 0, 150, 160 and 180 are similar to the results for Dolby Headphone. This model also has a tendency for 150 degrees being answered for many directions. Stimuli for 20 to 90 degrees end up in the 60 to 110 range. This is a range where Dolby Headphones have few answers for any direction. The head shadow filter might be the cause of this, lowering the signal for the left ear by too much, compared to the direction it should represent.

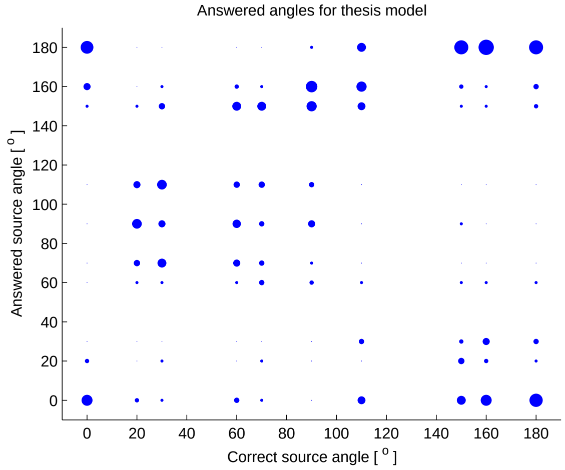
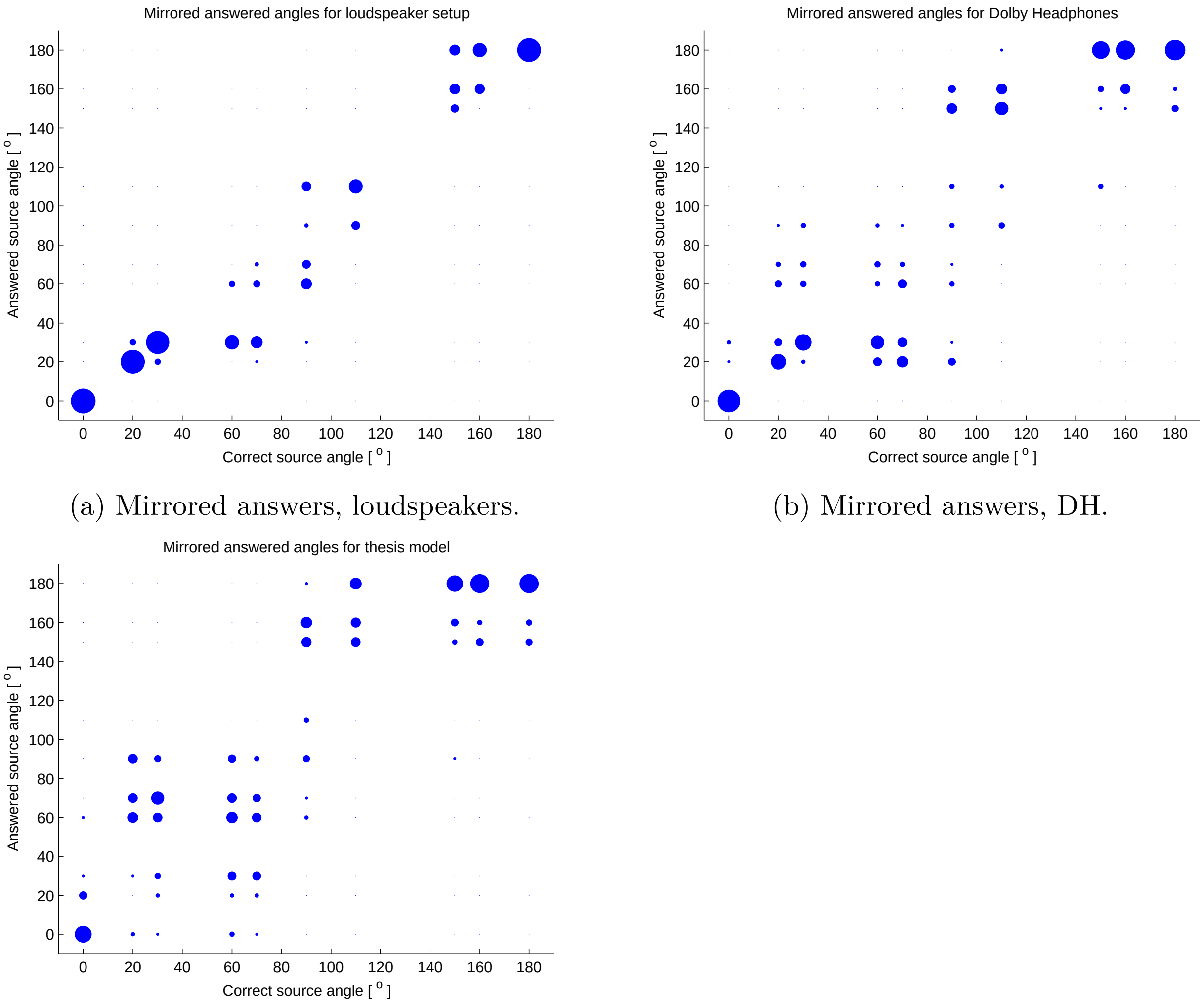


Figure 6.3: Raw answers, thesis model.

6.2 The mirrored results

The mirrored results for the three test types are shown in figure 6.4a, 6.4b and 6.4c. The correction is, as mentioned, very strict. Mirroring every front-back error around 90 degrees. Table 5.1 shows how many of these errors occurred for each test type. A surprise was that there actually were these kind of errors for the loudspeaker setup. Almost 5 % of the answered had to be mirrored. If one looks at figure 6.1 one can see that there were test subjects that answered 0 degrees, when the correct direction was 180 degrees. This indicates that even with our own unaffected hearing, we can make these kinds of errors. There are also a few errors for the 150 and 160 directions. There are no mirroring for the frontal directions for the loudspeaker setup. This shows that our hearing is most accurate for the frontal plane, which is well-known. The amount of mirroring for the Dolby Headphone test is about 1/3, which is quite a lot. If one looks at the plot of the mirrored answers in figure 6.4b. One can see a dominance of the answers of 20, 30 and 150 degrees for the middle section. The back directions of 150, 160 and 180 degrees are almost equally answered. These kinds of groupings make it difficult to create compensation alterations in the models, since many source directions get perceived as the same, while some directions are not really perceived at all. The thesis model has mirroring for over half of the answers. This is a large amount of errors and represents an increase compared to Dolby Headphone. By looking at the plot in figure 6.4c, one can see the same type of grouping effect as seen for Dolby Headphone. The difference is that even more directions are grouped together. The mirrored answers are divided into three sub-directions; 0 degrees, 60-70 degrees and 150-180 degrees. Most of the mirrored answers are in one of these sub-directions, making any potential attempt of compensation virtually impossible. The correction is based on the assumption that a visual feedback would eliminate these errors completely. It is not likely that visual feedback would be able to correct all the errors, when the amount of errors are more than 50%, such as for the thesis model. Hearing sound from the back, when the visual feedback tells you that it is actually in the front, would be quite annoying.



(a) Mirrored answers, loudspeakers.

(b) Mirrored answers, DH.

(c) Mirrored answers, thesis model.

Figure 6.4: Mirrored results for the three test types.

6.3 Mean and variation of results

Figure 6.5 shows how the mean answer varies with source direction. For loudspeaker setup, it is fairly straight. For Dolby Headphone and thesis model, it resembles two connected arcs, connected at 70 degrees. This arcing can come from the task of converting a virtual source, that might appear somewhere on a straight imaginary line inside the subjects' head passing through both ears, to the half circle of valid answers on the interface. This effect looks like it is stronger with the thesis model, than Dolby Headphone. This is supported by the various comments made by the test subjects. Many commented that the sources for the thesis model appeared inside the head, while Dolby Headphone made the source appear more or less outside the head. Most comments also stated that it was generally easier to choose a source direction with Dolby Headphone, rather than with the thesis model. One can see from figure 6.5 that the shaded 95% confidence interval areas rarely overlap. By looking at the total mean and confidence interval values in table 5.2 one can see that none of the 95% confidence intervals for the three test types overlap at all. This means that there is a significant difference between these three test types. By looking at the means, one can see that the loudspeaker setup and Dolby Headphone have almost the same mean value, only different sign. While the thesis model has over triple the mean value compared to Dolby Headphone, mean error is not everything.

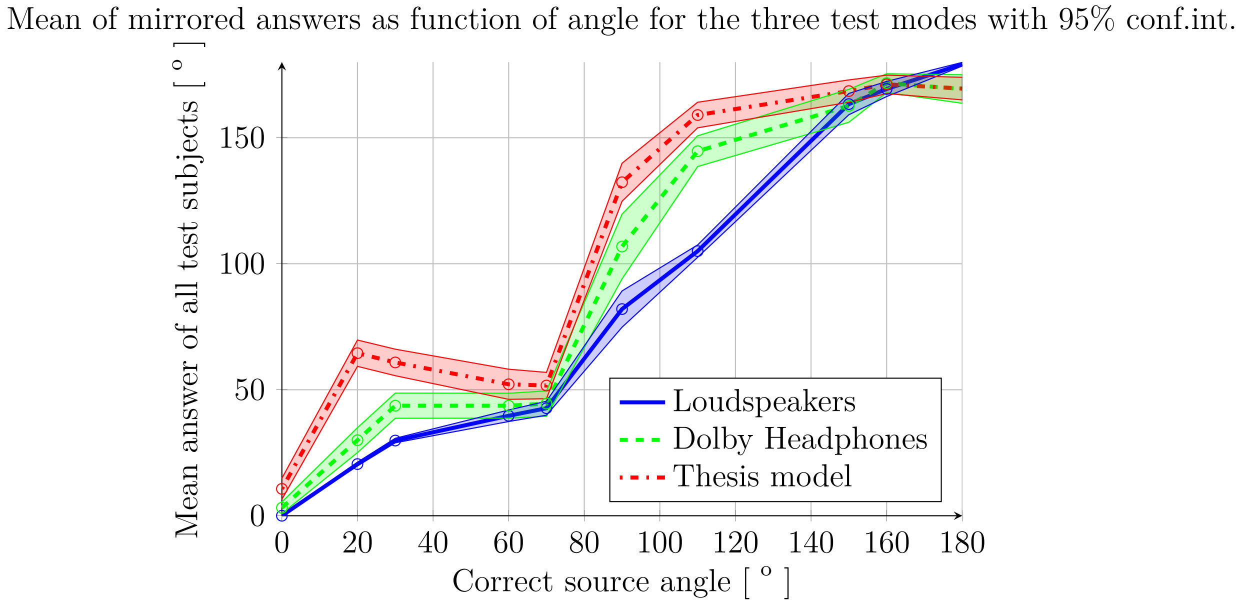


Figure 6.5: Mean of mirrored answers, with 95% confidence interval for the three test types.

If one looks at the standard deviation for the three test, one can clearly see that the loudspeaker setup provides the lowest deviation, while Dolby Headphone and the thesis model have essentially the same standard deviation. Figure 6.6 shows that the thesis model has roughly the same deviation for all directions. While both Dolby Headphones and the loudspeaker setup peak comparably more at 90 degrees, and vary more over direction than the thesis model. The standard deviation for Dolby Headphone is almost double the deviation for the thesis model or the loudspeaker setup at that direction. This shows that Dolby Headphones have a serious weakness for that direction. Keeping in mind that this is the standard deviation for the mirrored results, which will provide better results for all directions except 90 degrees, which is unchanged. This means that the deviations for the thesis model and the loudspeaker also are unchanged, so the comparability is valid.

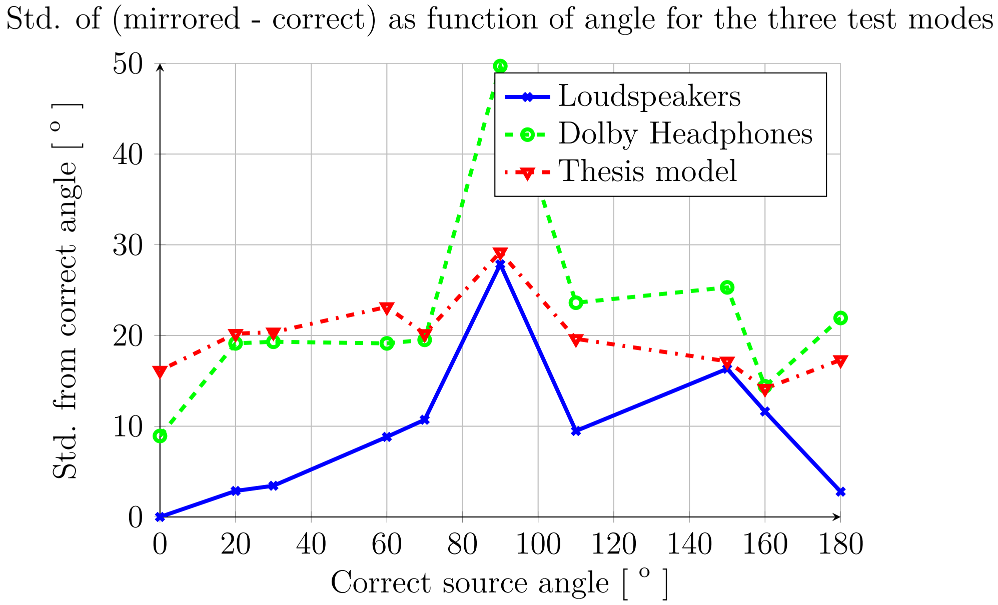


Figure 6.6: Standard deviation of mirrored answers for the three test types.

When it comes to choosing between a "consistent" deviation, such as for the thesis model, or a varying deviation, such as for the loudspeaker setup or Dolby, it might boil down to personal preference. Both the loudspeaker setup and Dolby Headphone have their lowest deviations for 0 degrees, which is the center channel. The center channel provides the speech track in movies. Locating the speech to either to one of the sides might become annoying, so a low deviation for 0 degrees might be wanted. Consistency might be more pleasant when it comes to gaming.

6.4 Subject consistency

Figure 6.7 shows what could be translated to answer consistency of the test subjects for the three test types. All but one test subject have a deviation in their answering of 5 to 10 degrees for the loudspeaker setup. The exception is test subject number 7, with over 30 degrees deviation for the loudspeaker setup. If one looks at the measured data for test subject 7 in appendix B, one can see that this test subject has an abnormal head shape and ear placement. This might be the reason for the deviation from the rest in the loudspeaker setup test. Test subjects like number 7 might be considered removed from the result calculation. This is not done here for two reasons. Firstly, because the test subject answered pretty average for the two headphone tests. Secondly, because this experiment and thesis is built around the concept of commercially available hardware, software and content presented to the average consumer. It is not less likely that test subject number 7 is a potential consumer of such a product. The great variation of deviation for the headphone tests might indicate their difficulty of creating distinctive directions, even if they are wrong directions as well. This makes it harder for the test subject to decide when answering. This can cause the test subject to rely more on guessing, which may indicate that the headphone models provide a more or less spatial illusion and less spatial information.

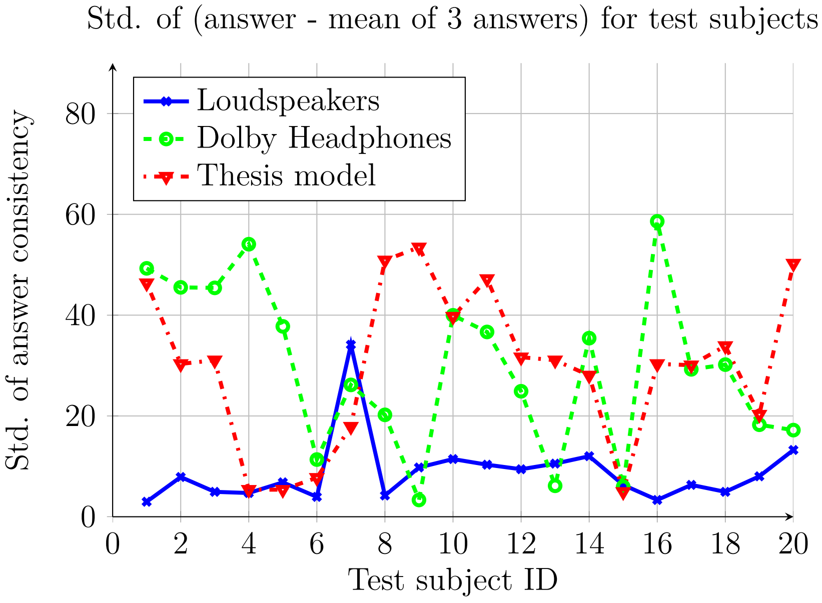


Figure 6.7: Standard deviation for each test subject. Mirrored answer minus the mean of the three answers for that source direction, for each test type.

6.5 Minimum results for a surround sound model

A surround sound setup or model should be able to deliver a minimum spatial unique locatable resolution. Figure 6.8 illustrates what this minimum spatial resolution could be. This resolution is in the horizontal plane only. A similar partition could be made for elevation as well. The horizontal plane should be mastered before including elevation. This horizontal resolution is 45 degrees. Loudspeakers, or virtual loudspeakers, should be placed at the center of each partition; 0, 180, ± 45, ± 90 and ± 135 degrees. If these would be localized with 90-95% certainty within the border of their partition, then we would have a working surround sound model that would actually live up to its name; surround sound. If one looks at the mean, including the 95% confidence span, localization direction in figure 6.9 (same figure as 5.2 but with the mentioned partition limits) at the aforementioned directions. One can see that all three test types manage to stay within their boundaries for 0, 45 and 180 degrees. Only the loudspeaker setup lies within the limits for 90 and 135 degrees. Dolby Headphone exceeds the limits, and the thesis model misses the limits completely. Figure 6.9 shows the mirrored results. The mirrored results have been used for most of the analysis in this thesis. Figure 6.10 shows how the raw results map out compared to the mentioned limits. The loudspeaker lies almost within the limits for all the directions, except a slight overshoot on 180 degrees. Dolby Headphones and the thesis model basically fails all across the board. Figure 6.10(NB! Read figure caption) shows that the sound for Dolby Headphone and the thesis model is localized within 50-150 degrees, which can pretty much be described as “to the right”. This is not really a localization result, since an ordinary stereo headphone would produce an equal result, but it shows how low these models perform in general. This poor performance is partially due to great amount of front-back confusion. Visual feedback will probably not be able to correct all these errors, and the actual result would be some kind of interpolation between figure 6.9 and 6.10 for Dolby Headphone and the thesis model.

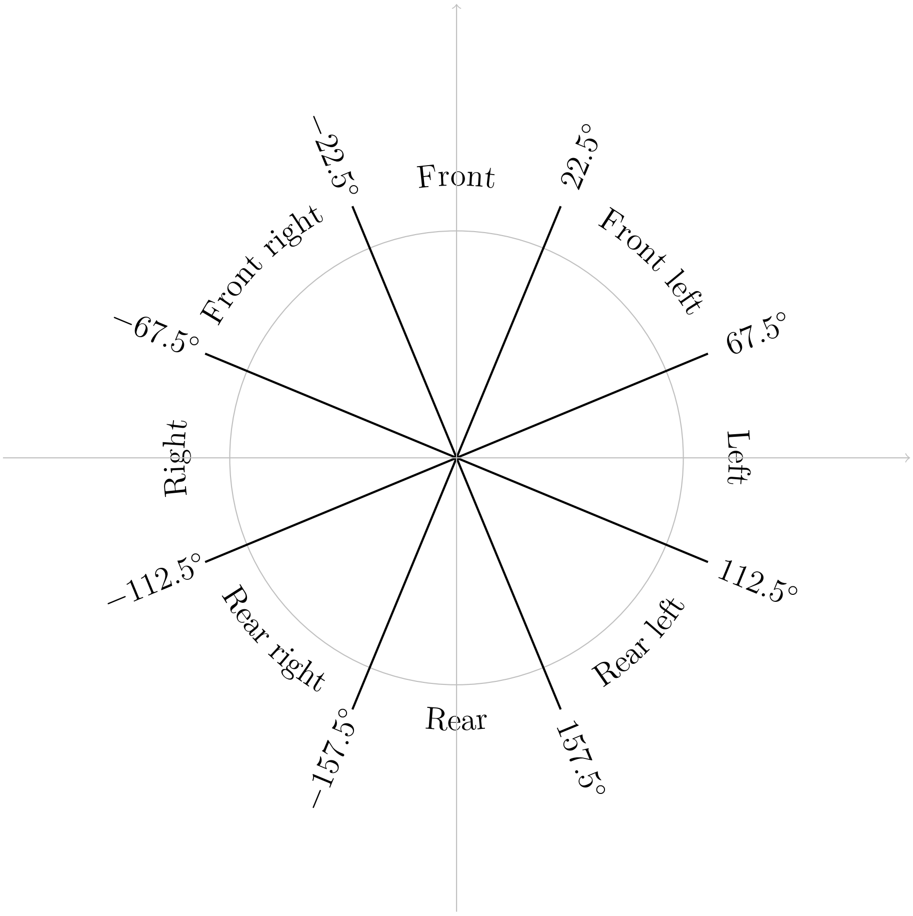


Figure 6.8: Illustration of minimum spatial resolution of what a surround sound setup or model should be able to produce.

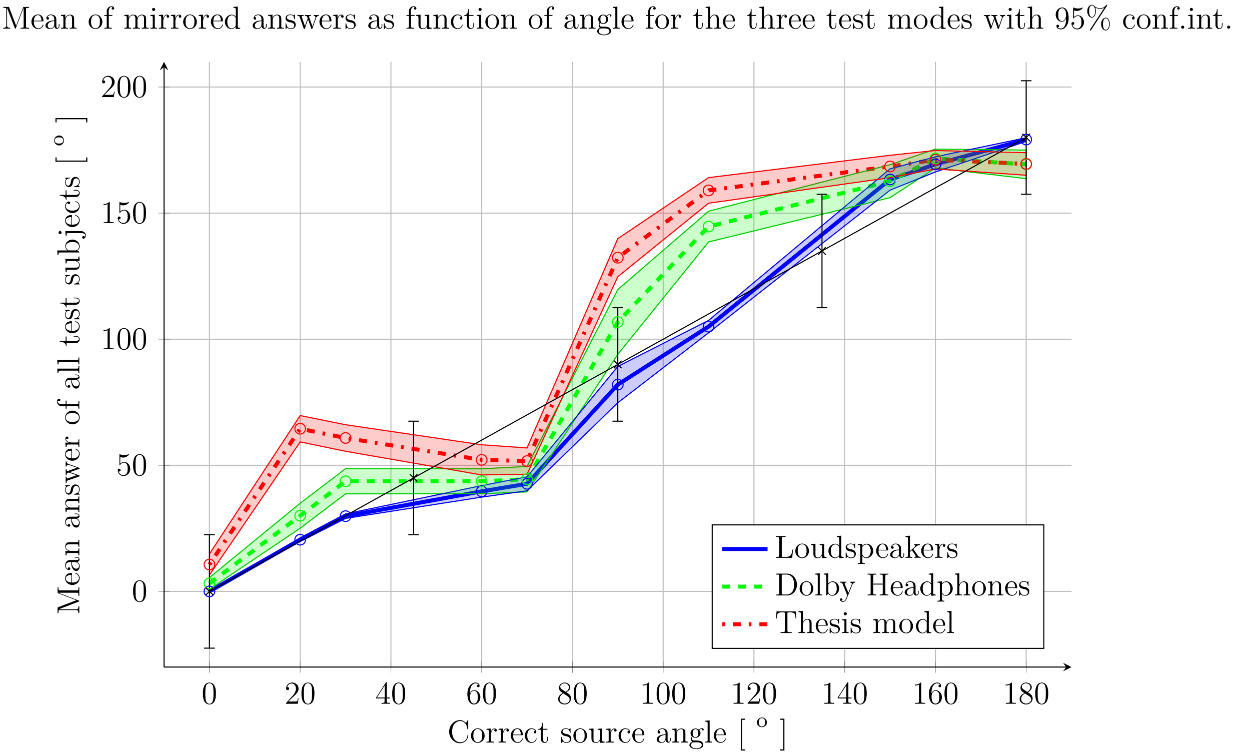


Figure 6.9: Mean of mirrored answers with 95% confidence interval. Bar shows the limits corresponding to the spatial partitioning of figure 6.8.

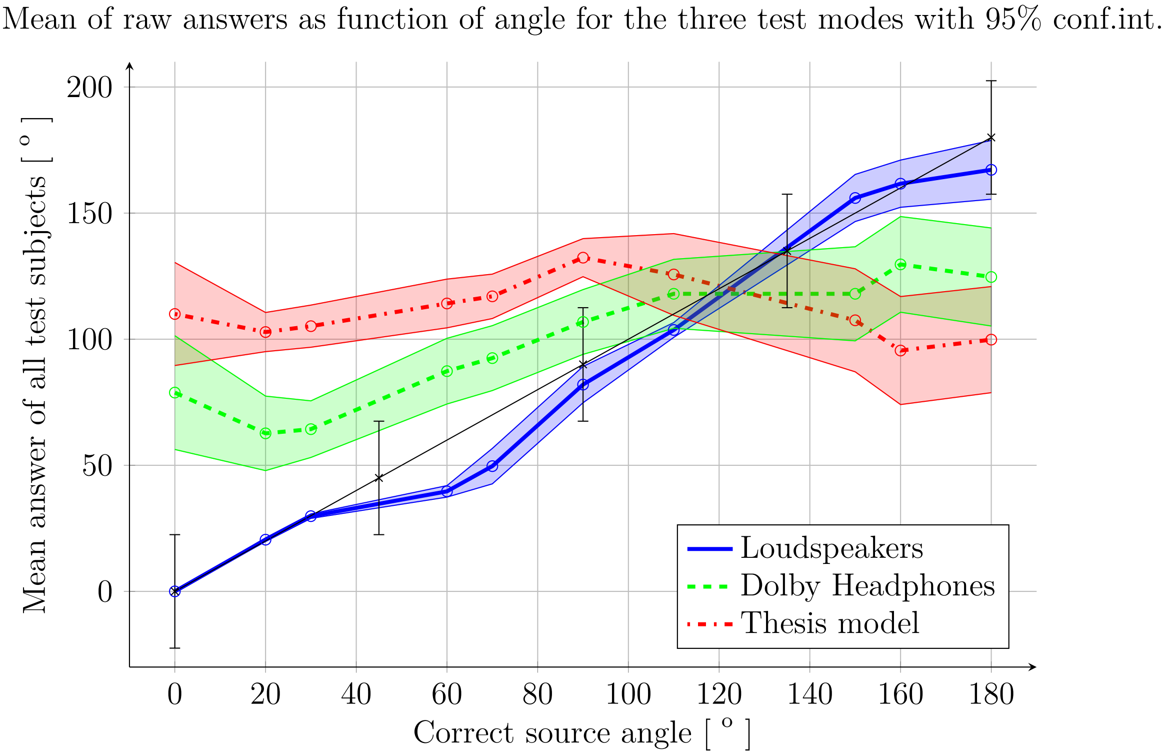
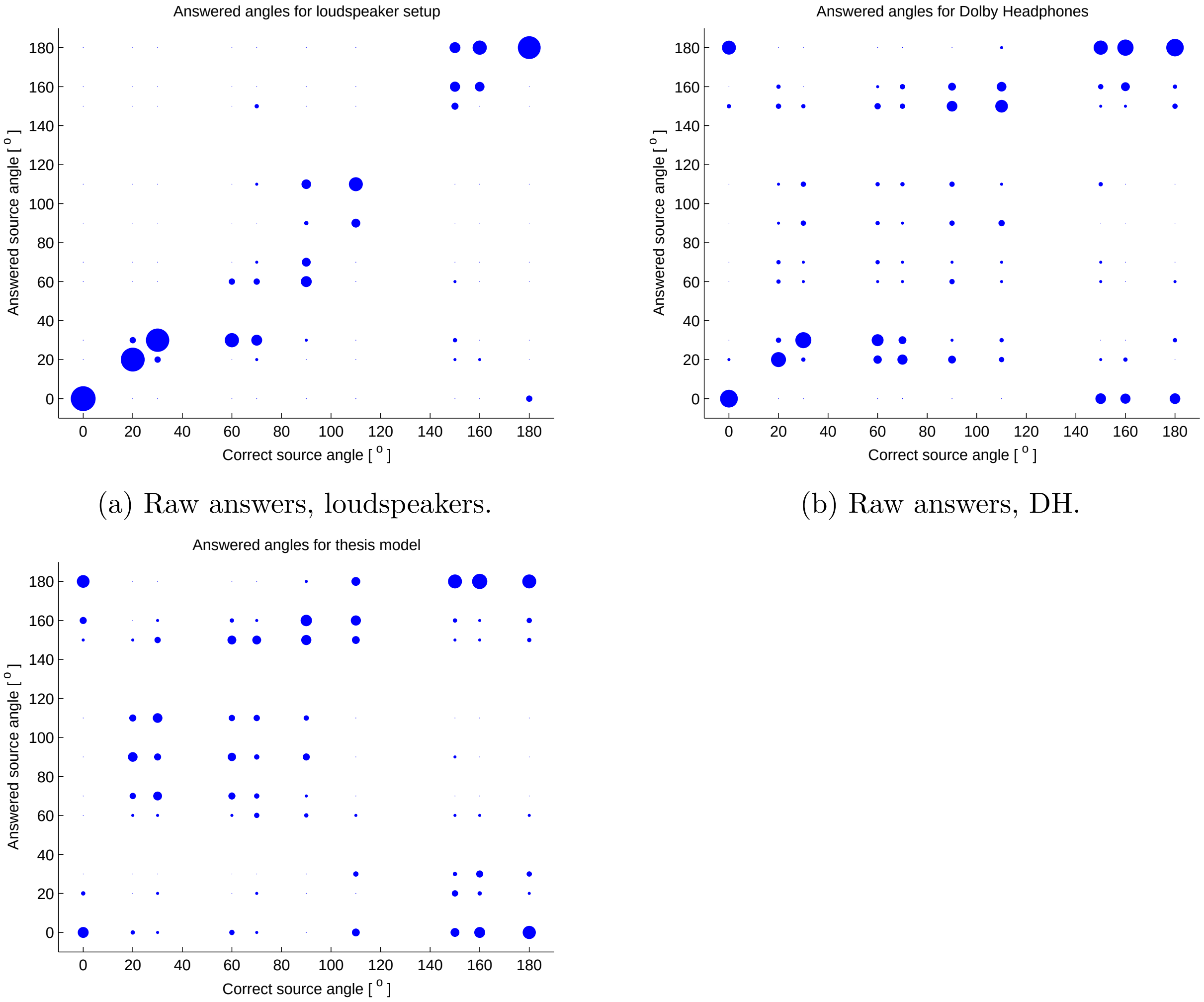


Figure 6.10: Mean of raw answers with 95% confidence interval. Bar shows the limits corresponding to the spatial partitioning of figure 6.8. This statistically analysis is not really valid, since it assumes Gaussian-like distribution. By looking at figures 6.11a, 6.11b and 6.11c one can see that the data is not distributed in that fashion.


Answered angles for loudspeaker setup

Answered angles for Dolby Headphones

Answered angles for thesis model

(a) Raw answers, loudspeakers.

(b) Raw answers, DH.

(c) Raw answers, thesis model.

Figure 6.11: Raw results for the three test types.

Chapter 7

Conclusion

The statistical data calculated from the results, showed that the commercial model, in this case Dolby Headphone, had a significant lower mean value for the difference between the mirrored answer and the correct answer. The standard deviation was more or less the same for the two headphone models. Does this confirm or disprove the hypothesis? The hypothesis at test was:

Hypothesis:

> The adaptable model will perform as well, or better, than the commercial model in localization virtual sound sources. In other words, the adaptable model will give a smaller, or equal, error compared to the error of the commercial model.

Since the 95% confidence intervals for the thesis model or Dolby Headphones did not overlap, i.e. there is a significant difference between these two models, one can safely conclude that Dolby Headphones produced a smaller error than the thesis model. Hence the hypothesis has to be disproved.

The results from the three test types all share the basic shape of the results. The only feature they have in common is the intensity panning between a few discrete positions. Intensity panning have clear and known weaknesses for the 5.1 loudspeaker setup. Results show that this weakness is transferred to the headphone models, and made progressively worse by doing so.

A desired minimum spatial resolution for surround sound has been presented. Both Dolby Headphone and the thesis model fail to fulfil this minimum resolution. The actual results are that these models produce a directional accuracy comparable to intensity panning of a stereo signal over headphones. The real conclusion, which was not the research goal, is that neither Dolby Headphones nor the thesis model can produce accurate surround sound. A few directions can at best be localized with the correction help of visual feedback.

Chapter 8

Further work

This chapter will mention, and briefly discuss, how the adaptable binaural surround sound model could be modified and how the experiment could be changed to highlight other aspects of these models. In closing there will be some thoughts on the future of surround sound in headphones.

- The general comment from the listening tests was that it was easier to pick a direction to answer with Dolby Headphone. This was due to that the sound appeared more out of the head, while the thesis model mostly appeared inside the head. This is probably because Dolby Headphone simulates an acoustic environment with reflections and reverberation. Similar features could be added to the thesis model.

- The results show that the thesis model localized very many directions further to the side than they should be. This might be due to a too strong amplification/attenuation by the head shadow filter. This is something that easily can be adjusted.

- The thesis model had over 50% front-back confusion errors. This points to that the spectral cues model chosen is not working as intended. A better model for spectral cues is needed.

- Another way to test the performance of the models would be to test moving sources. This could easily be done by changing the direction of the four noise bursts used to; target direction, plus five degrees, minus five degrees and then the target direction again.

- A quick reference of the key directions, such as 0, 90 and 180 degrees could be given to the test subject before the test. This would rule out mistakes due to unfamiliarity, that the test subject would not make if he/she owned the product, and had been using it for some time.

Even though the improvements mentioned above could give better results, the future of surround sound should take a different path. The solution lies in finding a quick, cheap and automated way to measure a persons own HRTFs, and offer it at the place they buy their equipment. Probably the quickest, and most potent solution would be a highly detailed 3D-model scan of the person. Then one could use numerical methods and/or ray tracing to calculate any HRTF needed. This is very computationally demanding, but computers get twice as fast every year, so this option is not that far into the future. This scanning would also fit into some sort of booth, and could take less than a minute. The computation of a HRTF database could be done somewhere else at a server park. 3D-scanning has been used to analyse pinna shapes, so this is already in motion.

Bibliography

[1] R. Duda, C. Avendando, and V. Algazi, "An adaptable ellipsoidal head model for the interaural time difference," in *IEEE International Conference on Acoustics, Speech, and Signal Processing*, vol. 2, pp. 965–968, 1999.

[2] E. Torick, "Highlights in the history of multichannel sound," *Jornal of the Audio Engineering Society*, vol. 46, pp. 27–31, February 1998.

[3] R. S. Woodworth and G. Schlosberg, *Experimental Psychology.* NY: Holt, Rinehard and Winston, 1962. pp. 349-361.

[4] G. Kuhn, "Model for the interaural time differences in the azimuthal plane," *Journal of the Acoustical Society of America*, vol. 62, pp. 157–167, 1977.

[5] V. R. Algazi, R. O. Duda, R. Duraiswami, N. A. Gumerov, and Z. Tang, "Approximating the head-related transfer function using simple geometric models of the head and torso," *Journal of the Acoustical Society of America*, vol. 112, pp. 2053–2064, November 2002.

[6] C. Brown and R. Duda, "An efficient HRTF model for 3-D sound," in *Applications of Signal Processing to Audio and Acoustics*, IEEE, October 1997.

[7] B. Gardner and K. Martin, "HRTF measurements of a KEMAR dummy-head microphone." `http://sound.media.mit.edu/resources/KEMAR.html`, 1994.

[8] V. Pulkki, "Virtual sound source positioning using vector base amplitude panning," *Journal of Audio Engineering Society*, vol. 45, pp. 458–466, June 1997.

[9] International Telecommunication Union, *Multichannel stereophonic sound system with and without accompanying picture*, July 2006.

[10] "Dolby 5.1 loudspeaker setup guide." `http://www.dolby.com/us/en/consumer/setup/connection-guide/home-theater-speaker-guide/select-config-5-1.html`.

[11] N. R. D. Corporation, "Ambisonic surround sound system." `http://www.imf-electronics.com/Home/imf/ambisonic`. Readable version found at URL.

[12] A. M. T. H. R. Noisternig, Markus; Sontacchi, "A 3D ambisonic based binaural sound reproduction system," in *Audio Engineering Society Conference: 24th International Conference: Multichannel Audio, The New Reality*, 6 2003.

[13] K. Brandenburg, "Wave field synthesis," in *3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, pp. 1–4, IEEE, 2009.

[14] B. Bauer, "Stereophonic earphones and binaural loudspeakers," *Jornal Of The Audio Engineering Society*, vol. 9, no. 2, pp. 148–151, 1961.

[15] B. Mikhaylov, "Bauer stereophonic-to-binaural dsp." `http://bs2b.sourceforge.net/`.

[16] D. S. McKeeg, Adam; McGrath, "Using auralization techniques to render 5.1 surround to binaural and transaural playback," in *Audio Engineering Society Convention 102*, 3 1997.

[17] "Dolby Headphone - how it works." `http://www.dolby.com/us/en/consumer/technology/home-theater/dolby-headphone.html#2-How-It-Works`.

[18] "Dolby Headphone - overview." `http://www.dolby.com/us/en/professional/technology/pc/dolby-headphone.html#1-Overview`.

[19] "Dolby Headphone - specifications." `http://www.dolby.com/us/en/professional/technology/pc/dolby-headphone.html#2-Specifications`.

[20] J. Borwick, ed., *Loudspeaker and Headphone Handbook.* Focal Press, third ed., 2001. Chapter 14.

[21] H. Møller, D. Hammershøi, C. B. Jensen, and M. F. Sørensen, "Transfer characteristics of headphones measured on human ears," *Audio Engineering Society*, vol. 43, pp. 203–217, April 1995.

[22] C. B. H. D. S. M. F. Møller, Henrik; Jensen, "Design criteria for headphones," *J. Audio Eng. Soc*, vol. 43, no. 4, pp. 218–232, 1995.

[23] D. Griesinger, "Binaural techniques for music reproduction," in *Audio Engineering Society Conference: 8th International Conference: The Sound of Audio*, 5 1990.

[24] V. Willert, J. Eggert, J. Adamy, R. Stahl, and E. Korner, "A probabilistic model for binaural sound localization," *Systems, Man, and Cybernetics, Part B: Cybernetics, IEEE Transactions on*, vol. 36, pp. 982 –994, oct. 2006.

[25] J. Huopaniemi and K. A. J. Riederer, "Measuring and modeling the effect of source distance in head-related transfer functions," *The Journal of the Acoustical Society of America*, vol. 103, no. 5, pp. 2988–2988, 1998.

[26] E. M. A. M. R. Begault, Durand R.; Wenzel, "Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source," *J. Audio Eng. Soc*, vol. 49, no. 10, pp. 904–916, 2001.

[27] E. Granier, M. Kleiner, B.-I. Dalenbäck, and P. Svensson, "Experimental auralization of car audio installations," in *Audio Engineering Society Convention 98*, 2 1995.

[28] M. F. J. C. B. H. D. Møller, Henrik; Sørensen, "Binaural technique: Do we need individual recordings?," *J. Audio Eng. Soc*, vol. 44, no. 6, pp. 451–469, 1996.

[29] K. Brørs, "Effekt av kombinert audio-visuell trening ved binaural gjengivelse," Master's thesis, Norwegian University of Science and Technology, June 2005. Thesis in Norwegian.

[30] C. Brown and R. Duda, "A structural model for binaural sound synthesis," *Speech and Audio Processing, IEEE Transactions on*, vol. 6, pp. 476 –488, sep 1998.

[31] C. D. R. O. Algazi, V. Ralph; Avendano, "Estimation of a spherical-head model from anthropometry," *J. Audio Eng. Soc*, vol. 49, no. 6, pp. 472–479, 2001.

[32] L. Rayleigh and A. Lodge, “On the acoustic shadow of a sphere. with an appendix, giving the values of legendre’s functions from p0 to p20 at intervals of 5 degrees,” *Philosophical Transactions of the Royal Society of London. Series A, Containing Papers of a Mathematical or Physical Character*, vol. 203, pp. pp. 87–110, 1904.

[33] J. Blauert, “Sound localization in the median plane,” *Acoustica*, vol. 22, pp. 205–213, 1969/1970.

[34] R. H. Y. So, N. M. Leung, A. B. Horner, J. Braasch, and K. L. Leung, “Effects of spectral manipulation on nonindividualized head-related transfer functions (HRTFs),” *The Journal of the Human Factors and Ergonomics Society*, pp. 271–283, June 2011.

[35] A. Barreto and N. Gupta, “Dynamic modeling of the pinna for audio spatialization,” tech. rep., Digital Signal Processing Laboratory, Florida International University, 2004.

[36] S. Spagnol, M. Geronazzo, and F. Avanzini, “Fitting pinna-related transfer functions to anthropometry for binaural sound rendering,” in *Multimedia Signal Processing (MMSP), 2010 IEEE International Workshop on*, pp. 194 –199, oct. 2010.

[37] K. Iida, M. Itoh, A. Itagaki, and M. Morimoto, “Median plane localization using a parametric model of the head-related transfer function based on spectral cues,” *Applied Acoustics*, vol. 68, no. 8, pp. 835 – 850, 2007.

[38] N. Gupta, A. Barreto, and M. Choudhury, “Modeling head-related transfer functions based on pinna anthropometry,” tech. rep., Digital Signal Processing Laboratory, Florida International University, 2004.

[39] Corsair®, “Vengance® Dolby 7.1 USB Gaming Headset.” `http://www.corsair.com/vengeance-1500-dolby-7-1-usb-gaming-headset.html`.

[40] P. Svensson and U. Reiter, “3D-sound/multimedia.” Lecture notes in the course TTT01 at Norwegian University of Science and Technology.

List of Figures

List of Tables

Appendix A

Description of MATLAB-functions on CD

A.1 Functions

[v1, v2, x00] = ellsection(A, x0, n, c)

Computes the ellipse that results from intersection between an ellipsoid with a plane. Ellipsoid is defined by (x - x0)'*A*(x - x0) = 1, and the plane is defined by n'*x = c. The return variable x00 is the 3 dimensional center of the ellipse, and v1 and v2 are vectors pointing in the major and minor axis of the ellipse. *Function supplied by R. Duda from the article [1].*

[v1, v2, x00] = elltan(A, x0, S)

Computes the ellipse that defines the points where rays from the source point S are tangent to the ellipsoid. Ellipsoid and return variables are the same as for the function *ellsection*, it also uses *ellsection* in its calculation. *Function supplied by R. Duda from the article [1].*

[L, Lan] = ellipsearc(a, b, t1, t2)

Computes the arc length, L, of an ellipse defined by x(t) = a*cos(t) and y(t) = b*sin(t), between angles t1 and t2 (given in radians). Based on a numerical method. Also calculates the approximate perimeter of the ellipse, *Lan*. *Copyright (c) 2010, Luc Masset All rights reserved.* Full license found on the CD.

mindist = mineldist(S, E, A, x0, gag)

Approximates the shortest distance from source position S to ear point E on an ellipsoid. Ellipsoid is defined in the same manner as in *ellesection* and *elltan. mineldist* uses *ellesection* and *ellipsearc* in its calculation. *ellipsearc* was replaced by the author of this thesis, since the old code used was discontinued in newer version of MATLAB. *Function supplied by R. Duda from the article [1].*

x = powernoise(alpha, dur, fs)

Generates $fs * dur$ samples of power law noise, with the spectrum of f^{-alpha}. *fs* is the sampling frequency, and *dur* is the duration in seconds. $alpha = 1$ will generate pink noise, and $alpha = 0$ will generate white noise. *(c) Max Little, 2008, Dale B. Dalrymple.* Explanatory readme-file for this function found on the CD.

nb = loadNoiseBurst(fs, ndur)

Loads the noise burst saved in *noiseburst.mat* if the file exists, if not the files is created with noise duration *ndur* in seconds, and with sampling frequency *fs*. Pink noise is generated by default.

[nb_sun nb_shadow] = headShadow(nb, T_sune, T_shadowe, w0, azimuth)
Takes input signal *nb* to be filtered, called nb since noise bursts were used in this thesis, along with ear angles *T_sune* and *T_shadowe* for the ear facing the source (sunny) and the ear in the shadow of the head. *w0* is the frequency based on head radii (as defined in equation 3.2), and *azimuth* is the source angle in the horizontal plane with reference to the center of the head. The return variables are the filtered versions of *nb* for both ears. Filtering according to 3.1.

g = intensityPan(azimuth, sound, filename, fs)
Creates a 6 channel wav-file, with filename *filename*, that will cause correct playback over a 5.1 loudspeaker setup. Intensity panning according to VBAP (see section 2.2.1), and panning between loudspeaker placed according to [9]. *fs* is the sampling frequency, and *g* is the gain matrix for the 6 channels. Since noise bursts were used in this thesis, the file generated features four repetitions of *sound* with a small pause in between.

y = directionalBoost(x, fs, az, thetaear)
Adds frequency boosting as described in section 3.3 to the signal *x*. *fs* is sampling frequency, *az* is source azimuth angle and *thetaear* is the ear position angle.

[shadow sunny] = generateBinaural(s, e, A, x0, w0, sound, azimuth, fs, c)
Generates the binaural signal for sunny ear and shadow ear. *c* is the speed of sound. This function uses the previous explained functions; *mineldist*, *headShadow* and *directionalBoost*. Output signal is normalized according to maximum value of the signal on the sunny side of the head.

[testSubjects index] = loadListenerDatabase()
Checks if the file *ListenerDatabase.mat* exists, if it does the list of Listener objects is return in *testSubjects* and the length of this list is returned in *index*. If the file does not exits, then it is created with *index* = 0 and an empty list. See definition of the Listener class below.

isDeleted = deleteListener(ii)
Deletes Listener number *ii* from the list of Listeners in *ListenerDatabase.mat*. See definition of the Listener class below.

Handle = plot_ci(X,Y,varargin)
Plots confidence intervals and patch between two confidence interval lines. The main line is specified by 1st column of matrix Y, whereas confidence intervals are determined by 2nd and 3rd columns. *varargin* allows setting parameters for the main line, the patch, and the confidence interval lines, such as line style, line width, color etc.. *Copyright (c) 2011, Zbigniew. All rights reserved.* Complete license found on CD.

matlab2tikz(varargin)
Creats tikz files for LaTeX from MATLAB figures. *Copyright (c) 2008–2012, Nico Schlömer. All rights reserved.* Complete license found on the CD.

[correctedAngle FBC] = frontBackConfusion(correctAngle, guessedAngle)
This function takes as input the correct source angle, and the test subject's guessed source angle. It then checks if there have been front-back confusion, i.e. the test subject have answered the source direction at the back, when it actually came from the front, or

visa versa. This is return in the variable *FBC* as 0 (no confusion) or 1 (confusion). If there have been a confusion, then the guessedAngle is mirrored into the right quadrant. The corrected angle is returned in *correctedAngle.* If there have been no confusion, then *correctedAngle* returns *guessedAngle.*

A.2 Class and class methods

Listener

id Anonymous representation of the test subject.

age Age of the test subject, used for general statistics about the test subjects.

gender Gender of the test subject, used for general statistics about the test subjects. Valid values M for male and F for female.

a1 Half head width in meters.

a2 Half head depth in meters.

a3 Half head height in meters.

eb Ear placement back from half depth, in meters.

ed Ear placement down from half height, in meters.

sitting_height Height of the ear canal entrance when the test subject was seated in the test area. Value in meters.

testOrder = 'LDM' Indicates the order of test modes between L = loudspeakers, D = Dolby Headphone and M = thesis model. This is the default value, the method *shufflePos(obj)* shuffles this test order.

sPosLS A vector containing the angles at test for the loudspeaker setup. This vector will get shuffled with the method *shufflePos(obj).*

sPosDolby Same as *sPosLS*, but will get shuffled different.

sPosModel Same as *sPosLS*, but will get shuffled different.

answersLS Vector for storing the test subjects answers for the loudspeaker setup. This vector will be compared to *sPosLS* during the analysis.

answerDolby Same as *answerLS*, but will be compared to *sPosDolby.*

answerModel Same as *answerLS*, but will be compared to *sPosModel.*

A = getA(obj)
Create the symmetric matrix A used defining the head in the ellipsoidal model. Uses class values $a1$, $a2$ and $a3$.

e = getShadowEar(obj,azimuth)
Checks which ear is the shadow ear from the source's *azimuth*, and assign e the right value. This e value is used for calculating the interaural time delay for the ellipsoidal head model.

rh_opt = optRadii(obj)
Calculates the optimal head radii from the values $a1$, $a2$ and $a3$ as described in equation 3.4.

getInput(obj)

Prompts for input in the MATLAB command window, and saves it in the class variables. This function calls *shufflePos()* at the end.

sufflePos(obj)

Shuffles *sPosLS*, *sPosDolby*, *sPosModel* and *testOrder* uncorrelated.

A.3 GUI function

varargout = FeedbackSystem2(varargin)

Combines all of the functionality described above in a GUIDE (MATLAB GUI Design Environment), with additional code to create program flow as intended.

A.4 Scripts

Statistics

Reads *ListenerDatabase.mat* and calculates the needed parameters, and plot them. NB! File is not general, and will only work properly with the attached *ListenerDatabase.mat*, the calculation however are transferable to a general purpose.

Appendix B

Test subject data

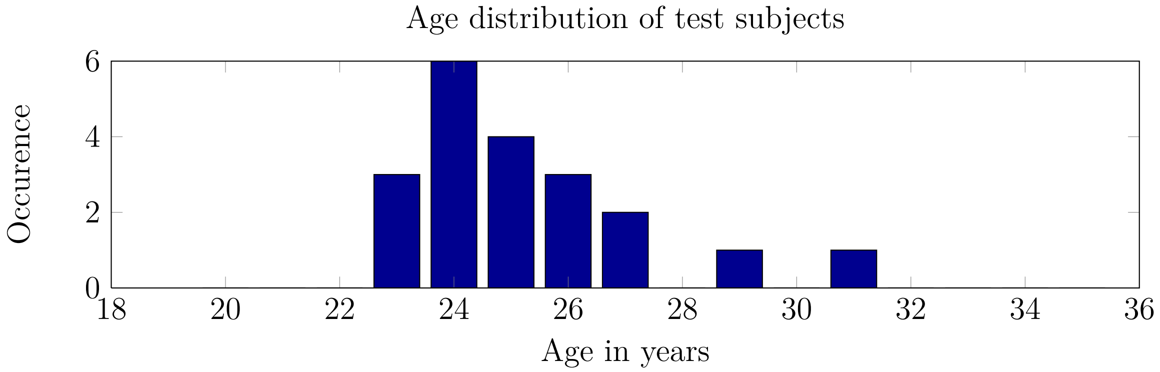


Figure B.1: Age distribution among test subjects.

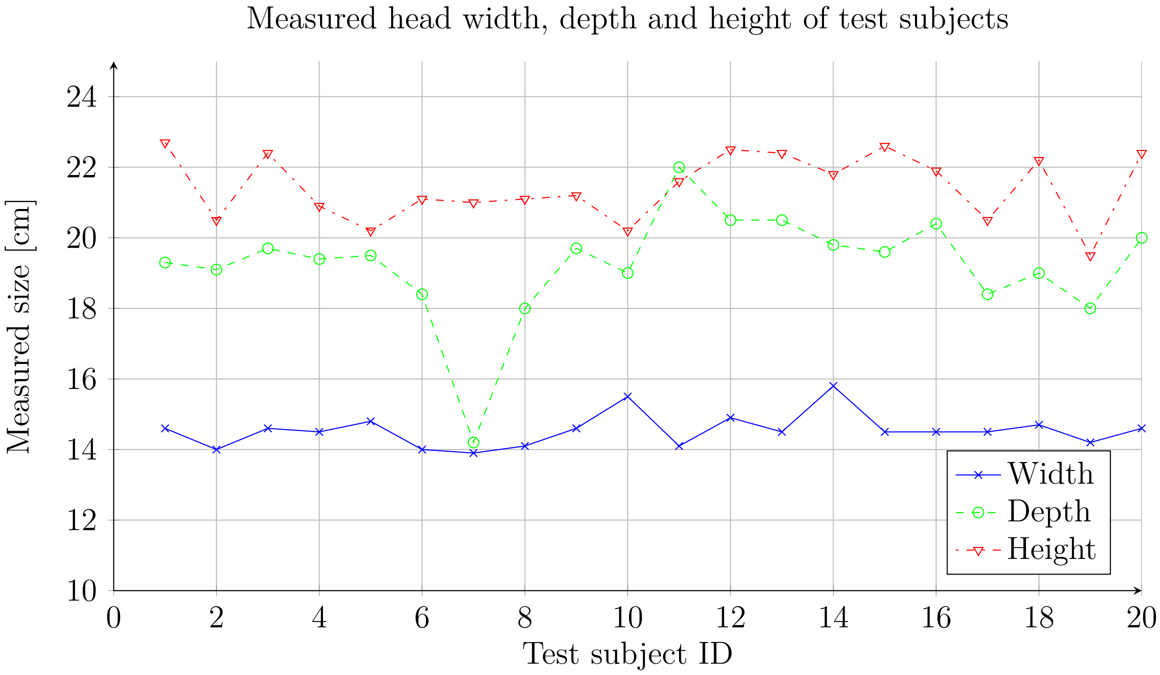


Figure B.2: Measured head width, depth and height for the 20 test subjects.

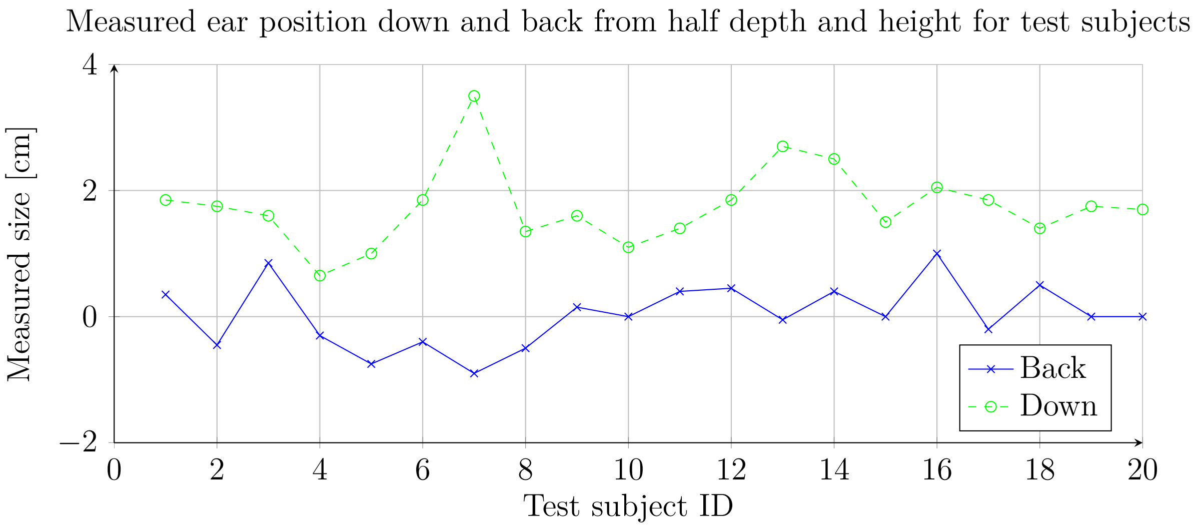


Figure B.3: Measured ear placement back and down for the 20 test subjects.

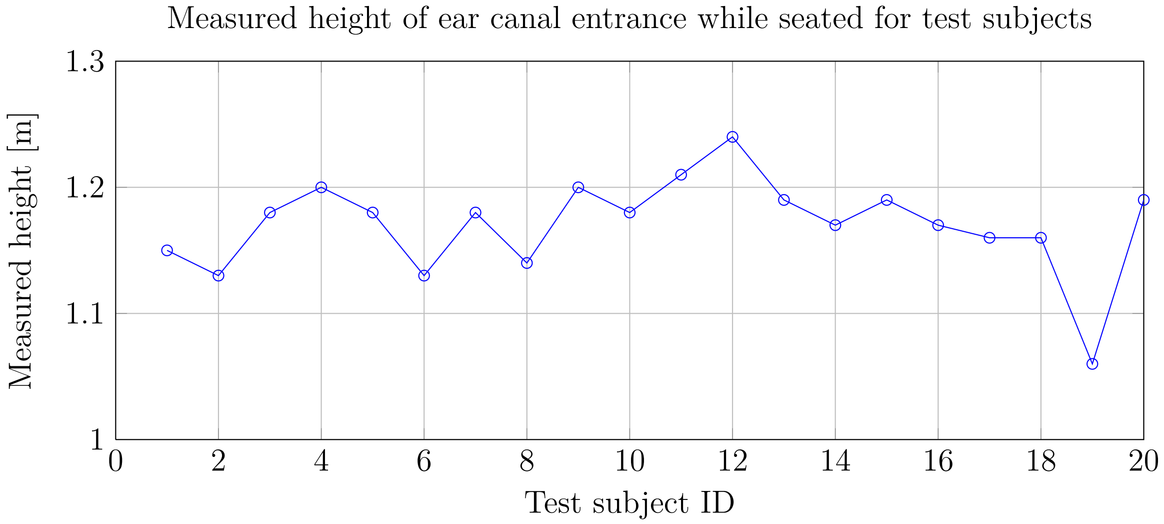


Figure B.4: Measured height of ear canal entrance while seated for test subjects.