Developing a Navigational HRTF-Based Audio Game for People with Visual Impairments

Brieger, Sebastian
brieger@kth.se

Degree project in: Media Technology
Supervisor: Hansen, Kjetil Falkenberg
Examiner: Li, Haibo

Project provider: Sound and Music Computing, KTH Royal Institute of Technology
Project Company: Funka Nu
Sound Hunter: Developing an HRTF-Based Audio Game for People with Visual Impairments

Abstract

In this thesis, I propose a framework for designing 3D-based audio-only games in which all navigation is based on perceiving the 3D-audio, as opposed to relying on other navigational aids or imagining the audio as being spatial, where additional sounds may be added later on in the development process. To test the framework, a game named Sound Hunter was developed in an iterative process together with both sighted and visually impaired participants in three focus groups, 8 usability tests, and a final evaluation. Sound Hunter currently features 20 levels progressively becoming more difficult, and relies on no stationary equipment. Instead, all navigation is based on the smartphone’s accelerometer data, where the only requirement is headphones to properly perceive the HRTF filtering, being delivered through the Pure Data external [earplug-], using a generalized HRTF and linear interpolation.

The results indicate that the suggested framework is a successful guidance tool when wanting to develop faster perception-based and action-filled 3D-audio games, and the learning curve for the navigation was approximately 15 minutes, after which the participants navigated with very high precision.

Furthermore, the results showed that there is a high need for audio-only games intended for visually impaired smartphone users, and that with only minor alterations to game menus and adjustments to the iPhone’s accelerometer function, both older and younger visually impaired smartphone users can navigate through 3D-audio environments using simple hand movements.

The results also strongly support that Sound Hunter may be used to train people’s spatial hearing in an entertaining way with full experimental control, as the participants felt that they focused more on their hearing and even trained their hearing while playing the game, and sounds were perceived as more externalized than lateralized. Also, the results strongly suggest that there are two main factors affecting the learning curve for adapting to a foreign HRTF during virtual interactive gaming experiences; the adaptation to the navigational controls, and the experience of front/back confusion, where control adaptation is promoted by having a strong default setting with customizable sensitivity, and the experience of front/back confusion can be greatly reduced by introducing complex distance-dependent meta-level communication in synthesized sounds. Using distance-dependent meta-level communication in the wrong way, however, can lead to illusions of false distance, making navigation impossible.

All of the participants would have recommended Sound Hunter to others, and they were very impressed by both the quality of the 3D rendering, and the way in which it could be used to navigate, where one of the participants, a blind expert audio-only game developer, also being highly experienced with audio-only games, claimed that Sound Hunter offered the best representation of 3D audio he had ever experienced in an audio-only game.

Keywords—Audio games, HRTF-synthesis, 3D-audio, game development, sonification, meta-level communication, visually impaired, blind, HCI.
Preface

In this project I have been working together with many different people. Not only have I achieved a better understanding of the current situation on audio game development, but also a developed understanding of the habits, behaviours, and needs of visually impaired people, and the importance of an appropriate engagement in these activities throughout the development of an audio game. Only then is it possible to develop something that will, step by step, become a real contribution to this user group. I am also sincerely thankful to my supervisor Kjetil Falkenberg Hansen at the Sound and Music Computing Group, who has given me support, motivation, and opportunities throughout the project, such as providing me the opportunity of writing a scientific article for the joint SMAC/SMC conference 2013 featuring Sound Hunter, also being accepted and published (Brieger 2013). Furthermore, I have had the pleasure of working together with the company Funka Nu, where I am especially thankful to Joakim Lundberg, offering me conference rooms and participants for the focus groups and usability tests, as well as helping me reach out to communities for visually impaired people. Special appreciations also go out to Philip at Blastbay Studios, with whom I have had the pleasure of working, also offering me expert opinions within the field of audio-only games intended for the visually impaired. Finally, I would like to thank Synskadades Riksförbund and Unga Synskadade Stockholm, through which I could get in contact with younger visually impaired people interested in audio games.
# Table of contents

1 Problem definition and purpose .................................................................1
  1.1 Research questions .............................................................................2
  1.2 Process overview ..................................................................................3

2 Theory ......................................................................................................4
  2.1 Sound ..................................................................................................4
    2.1.1 Sound waves .................................................................................4
    2.1.2 Spectrum and frequency components ...........................................6
    2.1.3 Digital audio signals ....................................................................7
    2.1.4 Sampling and quantization ..............................................................7
    2.1.5 Audio filters and effects ................................................................9
    2.1.6 Impulse responses .......................................................................10
    2.1.7 Convolution ...............................................................................10
  2.2 Human hearing ..................................................................................11
    2.2.1 Absolute thresholds for monaural and binaural sounds ...............11
    2.2.2 Hearing impairments ....................................................................12
  2.3 Directional hearing in humans ..............................................................13
    2.3.1 Common terms and concepts .........................................................13
    2.3.2 Cues for localization ..................................................................13
    2.3.3 Distance judgements ..................................................................17
    2.3.4 Other influences on our directional hearing ................................17
  2.4 Human vision ....................................................................................20
    2.4.1 Visual impairment ......................................................................20
  2.5 Spatial sound reproduction .................................................................21
    2.5.1 “Out-of-head experience” ..............................................................21
    2.5.2 Binaural sound and Head Related Transfer Function (HRTF) .........21
  2.6 Audio games .....................................................................................25
    2.6.1 Introduction .................................................................................25
    2.6.2 Audio games and the visually impaired ........................................25
    2.6.3 Audio-only games ......................................................................26
    2.6.4 Building audio-only games ..........................................................26
    2.6.5 Meta-level communication and sound characteristics ..................27
  2.7 3D-based audio-only games ...............................................................27
    2.7.1 Current problems with 3D-based audio-only games .................27

3 Proposed Framework ...........................................................................29
  3.1 Rethinking The Design of Navigational 3D-Based Audio-Only Games ..29

4 Method ....................................................................................................31
  4.1 Why use HCI methods? .....................................................................31
    4.1.1 Human Computer Interaction .....................................................31
    4.1.2 Mixed methods design ...............................................................31
    4.1.3 Iterative development ..................................................................32
5 Results

5.1 Part 1 – Focus groups: Developing Sound Hunter

5.1.1 Sound Hunter: The Original Game Idea

5.1.2 Focus group sessions

5.1.3 Participants

5.2 Focus group session 1

5.2.1 Qualitative data analysis

5.2.2 Results

5.2.3 Post-development

5.3 Focus group session 2

5.3.1 Qualitative data analysis

5.3.2 Results

5.3.3 Post-development

5.4 Focus group session 3

5.4.1 Qualitative data analysis

5.4.2 Results

5.4.3 Post-development

5.5 Part 2 – Testing and evaluating Sound Hunter

5.6 Usability tests

5.6.1 Participants

5.6.2 Procedure

5.6.3 Quantitative data analysis

5.6.4 Qualitative data analysis

5.7 Final Evaluation: Sound Hunter and the future

5.7.1 Participants

5.7.2 Qualitative data analysis

5.7.3 Results

6 Conclusions

6.1 Summarizing the results

6.2 Future research

6.3 Last words

7 REFERENCES
1 Problem definition and purpose

Even though the attention to game audio has increased over the years, the audio-only games today are not only very underdeveloped, but also extremely rare compared to mainstream audio-visual games. While they offer an increased spatial freedom due to the lacking graphical representation, this freedom can seldom be used, as almost all of the audio-only games are made for the PC, thus binding the player to stationary equipment. Furthermore, compared to the fast, action-filled audio-visual games, featuring realistic environments also accurately responding to the player’s navigational input, the tempo in audio-only games is usually slower, and instead of perceiving auditory spatial environments, the player has to rely more on their own imagination. A common argument among audio-only game developers is that an increased level of imagination is something positive, as it leads to an increased level of immersion, similar to when reading books. However, regardless of the level of realism in the game, the player will still always have to rely on their imagination, as the game world is not the real world. Therefore, I argue that the reason behind the success of audio-visual games, is most likely not due to an increased level of imagination, as the imagination will always be there. Instead, the success of audio-visual games seems to be that they have become increasingly realistic, and there is no reasonable argument as to why this should not be the case also for audio-only games.

Creating more realistic audio-only games may be accomplished by involving 3D-audio through Head-Related Transfer Funcion (HRTF) synthesis. However, 3D-based audio-only games having attempted this so far still have flaws. Not only are they even less common than audio-only games, but the 3D-audio is usually used as a complementary spatial effect, rather than being the foundation for perception-based 3D-audio navigation. For the 3D-based audio-only games focusing more on navigation, the problem instead seems to be their reliance on physical player movement, as well as additional stationary equipment, such as head-tracking devices or joysticks. Finally, in most audio-only games, little attention is to paid human sound perception, and the auditory environments are filled with too many sounds, either being HRTF-filtered, or being played back in stereo at a high volume, causing severe difficulties in localizing sounds.

The main target group for audio-only games is visually impaired people. Still, it is extremely rare for audio-only games to be developed together with the target group, as well as making use of HCI methods in the development process. Also, as the number of visually impaired iPhone users is rapidly increasing, but the number of audio-only games for smartphones can be counted on one hand, there seems to be not only a need for audio-only games intended for smartphone users, but also a high market potential for audio-only games for smartphones.

Yet another, more medical aspect as to why more audio-only games should be developed, is the fact that they may be used to train a person’s hearing, or teach the player to focus more on what they hear. Therefore, if truly navigational 3D-based audio-only games were to be developed, this would allow training the full scope of a person’s spatial hearing in virtual auditory environments, offering full experimental control.

In this thesis, the main purpose of developing Sound Hunter was not to develop an astonishing 3D-based audio-only game, but rather to examine the best way to make use of 3D audio for the purpose of navigation in similar games. When attempting to do this, it becomes highly important to develop a fundamental understanding of the properties of sounds and how they behave, both in real physical spaces, and during sound reproduction. Additionally, it becomes important to understand our natural capabilities and limitations when perceiving these sounds, as well as the auditory cues they create for our directional hearing.

However, as the development project was intended for visually impaired iPhone users, I also wanted to find out how the game could be optimized in order to meet the appropriate requirements and needs of this particular user group. I also wanted to compare the opinions of
Introduction

sighted people with those of visually impaired people, to find out whether or not sighted people regarded the game as being equally entertaining, as well as to find other differences between these two groups.

Finally, I wanted to find out whether or not the participants regarded Sound Hunter as training their hearing, and if they felt that the HRTF-based 3D-audio, provided by the means of studying human sound perception and implemented in a simple game, actually added something that more advanced commercial audio-only games today lack.

The three main research questions, as well as their individual sub-questions in connection to the same area of research, can be viewed in detail below.

1.1 Research questions

1. Research question 1

“Is it possible to build a navigational HRTF-based audio-only game where all navigation is based on perceiving 3D-audio cues, as opposed to using additional sounds to aid navigation?”

1.1. If so, can the game be built such that it does not rely on stationary equipment?
1.2. Can the game be built such that the player does not have to rely on physical movement (e.g. walking or turning)?
1.3. How accurately will participants be able to navigate by using only HRTF synthesis?
1.4. How long will the learning curves be for the navigation?
1.5. How does HRTF synthesis differ from stereo, and what are the main advantages and disadvantages of using each method when creating an audio-only game?
1.6. Should stereo and HRTF synthesis be combined when building HRTF-based audio-only games? If so, how?
1.7. What properties of the sounds can be adjusted in order for the game to become easier or more difficult without ruining the 3D-audio experience?
1.8. Can a design framework for developing HRTF-based audio-only games be created such that it applies to 3D-audio only games in general?

2. Research question 2

“Is there a need for HRTF-based audio-only games amongst visually impaired smartphone users?”

2.1. If so, how should the game be developed in order to be as compatible as possible with the way in which blind people use smartphones?
2.2. Does HRTF synthesis add something that is lacking in today’s audio-only games (e.g. directional hearing, distance perception)?
2.3. Will sighted participants find the game equally entertaining as visually impaired participants, or will there be a difference between these groups?
2.4. How may the game be developed further in order to become as entertaining as possible?

3. Research question 3

“Is it possible to use HRTF-based audio-only games to train a person’s hearing in an entertaining way?”

3.1. Will there be a difference between the learning curves for passive azimuth localization (only listening) and interactive azimuth localization (while gaming)?
3.1.1. Are research questions 1.4 and 3.1 related? In that case how?
3.2. Will participants experience sounds as being more externalized than lateralized?
3.3. Will participants find sound files and synthesized sounds as being equally entertaining? If so, how may these different types of sounds be used to train a person’s hearing?
1.2 Process overview
The research questions were answered through an extensive development process being presented below.

**Developing a framework** by studying a variety of existing types of audio games (e.g. audio-visual games, audio-only games, and 3D-based audio-only games) to create a framework for developing 3D-based audio-only games, as well as to come up with the initial game idea based on this framework.

**Developing the game: Sound Hunter** through three iterative focus groups, together with the development team consisting of two visually impaired participants, to gain qualitative feedback and perform various tests related to perfecting spatial perception, where each focus group was followed by post-game developments.

**Testing Sound Hunter** through a series of usability tests with eight participants, both sighted and visually impaired, including quantitative and qualitative data analyses.

**Evaluating Sound Hunter** with a young visually impaired participant, also being an expert in audio-only game development in order to evaluate the results, find ways of perfecting the existing game, as well as to discuss the types of future 3D-based audio-games that may be developed by following the framework and conclusions having been made throughout the entire process.
2 Theory

In order to understand human sound perception and spatial sound reproduction, one first has to understand what sound really is. I will therefore begin by presenting some of the physical properties of sound. Following this, I will discuss human hearing, human vision, and finally audio games.

2.1 Sound

When it comes to sound and human sound perception, there is a lot of theory to dig into. In this thesis, I will cover the parts that I find are the most essential to understand when developing an audio-based game. For more detailed information regarding the physical properties of sound waves, I recommend the book An Introduction to the Psychology of Hearing by Brian Moore. I will begin by giving a brief introduction to some of the basic concepts being important for a proper understanding of this report.

2.1.1 Sound waves

Sound can be explained as our perception of sound pressure variations at our eardrums. The range of sound pressure variations being audibly perceptible by humans is roughly between 20 and 20 000 variations (Hz) per second. Sounds are usually described as waveforms being shown on a two-dimensional diagram, where x corresponds to time and y corresponds to divergences from normal air-pressure.

In this report, a source of waves will be referred to as a sound object. Additional common terms for describing sound waves include:

- **Period/Wavelength** – Interval during which the sound wave is repeated (or the inverse of the frequency, see Figure 1)
- **Frequency** – Number of periods/variations per second, given in Hz (also referred to as temporal frequency, see Figure 1)
- **Amplitude/Sound level** – Maximum deviation from the nominal sound pressure level during each period, see Figure 1
- **Loudness** – The psychological correlate of amplitude, measured as subjective sensations from quiet to loud, usually being compared to the loudness of a 1kHz reference tone (Moore 2008)

---


**Sound**

Figure 1: Wavelength, frequency and amplitude in a sound wave (source: http://www.divediscover.whoi.edu/expedition12/hottopics/images/sound1-en.jpg)

Figure 2: Phase shift between two waves (source: http://ffden2.phys.uaf.edu/212_spring2011.web.dir/michael_hirte/Phase_shift.jpeg)

**Phase** – Relative position of the sound wave in time or space, see *Figure 2*

**Oscillation** – The amount that a sequence (of waveforms when talking about sounds) varies between extremes, see *Figure 3*
2.1.2 Spectrum and frequency components

According to the mathematical theory of Fourier series, periodic functions (in this case waveforms) can be explained as a sum of a (possibly infinite) set of simple oscillating functions (sine waves or cosine waves) (Pohlman 2010). This means that any sound wave can be explained as a combination of many different sine waves varying in frequency, amplitude and phase. As these sine waves can be seen as components of the sound wave, they are also usually referred to as frequency components.

When using a spectrogram (showing frequency and amplitude, i.e. the spectrum), a sound wave can be analysed in order to find all the frequency components within the sound. In the examples below, the results of a spectrum analysis are shown for a single sine wave at 50 Hz (Figure 4), as well as a more complex sound containing many sine waves (Figure 5).

---

2.1.3 Digital audio signals

Now that various qualities of sound waves have been explained, it is also necessary to get an understanding of how sound waves are treated in the digital domain, the reason for this being that this theory is highly connected to how the various game levels in Sound Hunter were constructed. For more detailed information about digital audio signals, the book *Principles of Digital Audio* by Ken Pohlman is recommended.

2.1.4 Sampling and quantization

In order for a physical sound wave signal to become a digital signal, it has to be represented as a numerical description, which is done by letting physical audio signals pass through an analog-to-digital converter (A/D), in which the audio signal is sampled and quantized.

Sampling is the process of reducing a continuous signal to a discrete signal. The *sampling frequency* is the amount of samples per second (given in Hz), where one sample corresponds to a set of values describing the signal at one point in time or space (Figure 6). When sampling an audio signal, the sampling frequency has to be twice as high as the highest frequency component in the audio signal (i.e. it has to be sampled at least twice per period), which is referred to as the Nyquist theorem (Pohlman 2010)³. If the audio signal is sampled correctly, the digital representation of the audio signal can be created *without any loss of information* (Pohlman 2010)⁴. For example, in a normal CD, the sampling frequency is 44.1kHz, meaning that the highest frequency component is 20.5kHz (half of the sampling frequency), which is higher than the highest audible frequency among humans (~20kHz). If, however, the sampling frequency is lower than twice per period of the highest frequency in the audio signal, aliasing will occur in the digital representation, meaning that higher frequency components will be represented as lower frequency components. This gives distortions and artifacts to the signal, as false frequencies are introduced. In Sound Hunter, all of the sounds were sampled at 44.1kHz.

Quantization is the discrete level representation of the waveform’s position, which can be done with varying accuracy. The quality of the sound is highly dependent on how the quantization is performed, as well as how many bits are used to represent the signal (e.g. when quantizing an audio signal with too few bits, this will result in the audio signal being perceived as cut into several bits), see Figure 7.

When it comes to the audio files used in Sound Hunter, it will be assumed that those gathered from external sources (i.e. websites offering free sounds) were quantized with a proper bit depth. The remaining audio files (recorded) were quantized at a bit depth of 16 bits per sample.
2.1.5 Audio filters and effects

Audio filters affect the phase or amplitude of the frequency components in a digital audio signal, which in turn affects the audio signal’s spectrum and alters its sound. The way in which a filter affects the phase and amplitude in an audio signal depends on the frequency, and is referred to as the filter’s *transfer function*. Transfer functions are usually shown in diagrams where the frequency represents x, and either the amplitude or phase represents y. Examples of common filters are:

**Low-pass filter:** Decreases or cuts of higher frequencies and lets lower frequencies pass through

**High-pass filter:** Decreases or cuts of lower frequencies and lets higher frequencies pass through

**Band-pass filter:** Decreases higher and lower frequencies within a certain frequency range

**Parametric equalizer (EQ):** Used to balance the frequency components in an audio signal. The parametric EQ has one or more sections making use of *second-order filter functions*, which are used to control three parameters (selection of the centre frequency, Q determining the sharpness of the bandwidth, and gain to boost or cut the centre frequency relative to the other frequencies outside the range of the bandwidth)\(^5\), see Figure 8.

![Figure 8: A parametric EQ where the first second-order filter function is activated as a high-pass filter. The three knobs (top to bottom) show the centre frequency, gain, and Q (source: screenshot from the music production software Ableton Live)](image)

Filters can also be created in other ways to obtain *audio effects*, for example by feeding the audio signal back into itself, by adjusting loudness and spectral content over time, or through *convolution* (explained below).

**ADSR Envelope generator:** Adjusts the loudness and spectral content over time, which can be used to simulate different instruments. An envelope generator has four main parameters\(^6\):

1. **Attack time:** The time from zero to peak loudness, when the midi key is pressed
2. **Decay time:** The time of rundown between the attack level and the sustain level
3. **Sustain level:** Level of the main sequence of the sound’s duration, until the midi key is released
4. **Release time:** The time from sustain level to zero when releasing the midi key

---


When feeding copies of the signal back into the original signal, a variety of effects can be created, for example:

**Echo:** One or several copies of the signal are added back into the original signal with a delay of typically 35 milliseconds or more.

**Reverb/Reverberation:** As opposed to simple echoes, reverberation makes use of very many delayed signals (up to several thousand echoes), which are added in quick succession (0.01-1 milliseconds between each other). Reverb is often used to simulate various acoustic environments (e.g. where the sound continues to decay even after the original sound can no longer be heard). This is done by varying the length of the decay (also referred to as *reverberation time*) as well as other variables, such as simulating *room size* or *surface types*.

**Chorus:** Short (~5-10 milliseconds) but constant delayed signals, often also *pitch shifted*, are added to the original signal in order to create voicing effects or to slightly smooth the signal.

What is important to realize here, is that all reflections are made up by delay, relative loudness (to the sound object), and frequency response, where the delay is determined by the *room size*, and the *relative loudness and frequency response* is determined by the *reflectivity and construction of the surfaces* (Pohlman 2010).

### 2.1.6 Impulse responses

A signal with high amplitude during a (shortest possible) period of time without repetition (or one sample in the digital domain) is referred to as an *impulse*. An important characteristic of an impulse signal is that when an impulse is sent through a spectrum, it shows equal magnitudes for all frequencies. Thus, when an impulse is passed through a filter, the spectrum of the signal output from the filter will be an exact copy of the filter’s transfer function. This resulting output signal being created by letting an impulse pass through a filter is referred to as the filter’s *impulse response* (Brieger & Göthner 2011). While some filters’ impulse responses are limited in time, others’ have infinite impulse responses. This is referred to as *FIR* (finite impulse response), and *IIR* (infinite impulse response) (Pohlman 2010).

The filters used in this report are linear and *time-invariant*, meaning that impulses varying in amplitude create equal impulse responses but with amplitudes corresponding to those from the impulse signals.

By using impulse responses and *convolution*, almost any type of linear filter can be recreated.

### 2.1.7 Convolution

As mentioned above, every filter will always have a certain impulse response that contains all the information necessary to completely reproduce its transfer function. As every linear filter produces equal impulse responses for all impulses (except for the amplitude variations in the input impulses), it becomes quite easy to know how a filter’s impulse response will look like when being fed with an impulse. By looking at a finite signal and analysing every sample in that signal as being individual impulses varying in amplitude and phase, it becomes possible to calculate the entire output signal simply by combining every possible impulse response from the input signal (Smith 1997). Therefore, if a filter’s impulse response is known (and finite), it becomes possible to recreate its transfer function without any operations in the frequency domain. This technique is called *convolution* (Smith 1997).

---

2.2 Human hearing

Human sound perception is very complex, and includes not only the physics behind sounds and physiological mechanisms in our bodies, but also our sensations and psychological connections to the sounds we hear in our daily lives. It is therefore a psychophysical study, and in this thesis, the main focus is on our ability to determine the direction and distance to a sound object. However, as our natural hearing limitations also connect to this thesis, I will begin by presenting our natural frequency thresholds, and how they are affected when we grow older or our hearing gets damaged.

2.2.1 Absolute thresholds for monaural and binaural sounds

The absolute threshold for a sound is the minimum detectable level of that sound in the absence of other sounds, and is measured in two ways (Moore 2008). The first method determines the minimum audible pressure (MAP) close to the eardrums, where sounds are delivered through headphones (i.e. monaural listening).

The second method, minimum audible field (MAF), uses a speaker to deliver the sounds in a large echoic chamber, where the measurements are made in the space previously occupying the centre of the listener’s head (i.e. binaural listening, as it accounts for the shape of the listener’s head and pinnae). Estimates of MAP and MAF measurements are shown in Figure 9.

As shown in the figure, the lowest frequencies (< 100 Hz) and the highest frequencies (> 15 kHz) have much higher detection thresholds, and are therefore more difficult to perceive when being played back at the same sound level. However, even if an individual has a threshold of as much as 20 dB above or below the mean for a specific frequency, this may still be considered as

---

normal. Furthermore, the difference between monaural and binaural listening is on average 2 dB, where binaural sounds are more difficult to detect than monaural sounds, and the most significant differences appear around 4 kHz (Moore 2008)\(^{14}\).

For the game level construction in Sound Hunter, the absolute threshold for different frequencies was therefore taken into consideration in order to create levels being either easier, or more difficult to perceive due to their frequency components.

### 2.2.2 Hearing impairments

When we grow older, our hearing gets worse. This is typically noticeable in the higher frequency domain, as the hair cells responding to quicker air pressure variations stiffen and die (referred to as sensorineural hearing loss (Moore 2008)\(^{15}\)), which for example makes it more difficult for us to perceive certain consonants in speech including high-pitched noise (e.g. S, F, T)\(^{16}\). However, our hearing also gets worse for lower frequencies, where frequencies under 1 kHz have more of a linear decrease, as opposed to the exponential decrease patterns for the higher frequency areas (see Figure 10). Hearing loss can also occur from overexposure to sounds either being very loud, or being exposed to quieter sounds over a longer period of time, and the treatment (e.g. hearing aids, sound therapy, surgery) depends on the type of hearing loss (i.e. conductive or sensorineural hearing loss)\(^{15,16}\).


Hearing loss can also appear in one ear only. Interestingly, however, is that the auditory system in various cases seems to be able to compensate for this loss. A particular phenomenon being relevant for this thesis is that of loudness recruitment, which occurs when there is a rapid growth in the sensation of loudness in the damaged ear when being presented with the same sound to the healthy ear, where the increase may be as much as 60dB (Moore 2008)\(^{17}\). The reason why this is relevant here is that people who suffer from hearing loss in one ear may still be able to play Sound Hunter, as long as there is a healthy ear with which the damaged ear can “catch up”.

---


\(^{16}\) High Frequency Hearing Loss Information, Treatment, causes and symptoms, http://www.hearinglosspill.com/high-frequency-hearing-loss/

2.3 Directional hearing in humans

Before trying to understand the various techniques used to simulate sounds as if they are positioned in certain positions or directions in space, it is important to properly understand our human hearing capabilities when it comes to localizing sounds. The ability to determine direction and distance to a sound source is limited among humans, and also varies depending on different factors, such as the angle of the sound source, the sound’s frequency, intensity, as well as the type of sound.

2.3.1 Common terms and concepts

2.3.1.1 Localization

The term localization refers to judgements of the direction and distance of a sound source (Moore 2008)\(^\text{18}\). Thus, the human ability of accomplishing this is referred to as the ability to localize a sound object.

2.3.1.2 Direction judgements

Localizing a sound object’s direction can be done in three dimensions (or planes), all of which form a coordinate system around the centre of the listener’s head position. The horizontal plane refers to sound objects being located to the left, right, in front and behind the listener, the frontal plane refers to sound objects being located in front, behind, above or below the listener, and the median plane refers to sound objects being located above, below, to the left or right of the listener.

By considering this coordinate system, the direction of a sound source can be described by using two angles, the azimuth angle (describing the sound object’s position around the listener’s head) and the elevation angle (describing how high or low the sound object’s position is positioned in relation to the listener’s head).

2.3.2 Cues for localization

In order to aid localization, humans make use of various cues related to the information in sound objects. Most of these cues work by comparing the information being delivered to the right ear to that of the left ear. Humans do, however, also have slight abilities in localizing sounds using only one ear (Moore 2008).

2.3.2.1 Interaural level differences

Interaural level differences (ILD’s), mostly affect our ability to localize a sound object left and right, and occur mainly when an incoming sound from the side of the head reaches one ear directly, while at the same time passing through and diffracting around the head in order to reach the other ear, creating an acoustic shadow (Heeger 2006). Due to their longer wavelengths, lower frequencies diffract more easily around an object the size of a human head, and therefore the of ILD’s become insignificant under 500Hz, see Figure 11.

---

Directional hearing in humans

Figure 11: Interaural intensity differences become noticeable for frequencies over 500Hz, when they create an acoustic shadow (source: http://www.cns.nyu.edu/~david/courses/perception/lecturenotes/localization/localization.html).

During Sound Hunter’s level construction, certain levels made use of frequencies lower than 500Hz, thus eliminating the player’s possibility of using ILD cues as navigational aid (i.e. when approaching the sound object, the overall sound level increases, whereas the levels in each ear change equally, see the section on Planning Sound Hunter’s game levels).

2.3.2.2 Interaural time differences

The next localization cue, interaural time difference (ITD), has to do with the time at which the signal from the sound object reaches each ear. If the distances from the sound object to each ear are not exactly the same, there will be a time delay at one of the ears, causing the ITD (Brieger & Göthner 2011), (Heeger 2006). Every possible position in a given space leads to a specific ITD, where the lowest difference is 0 (the distances from each ear to the sound object are equal), and the highest difference is around 690 µs (when the sound object is located straight to the right or to the left of the listener with no elevation) (Moore 2008)\(^1\).

In theory, the amount of positions in space giving equal interaural time differences is infinite, which is usually illustrated by marking the surface they create in relation to the listener’s head (Figure 12). The phenomenon has been named the cone of confusion, due to the cone-shaped surface created by the different positions. Because of the cone of confusion, it becomes difficult to localize the sound object by using interaural time differences as the only cue for localization (Moore 2008)\(^2\). In this thesis, the focus is mainly on the azimuth angle, and the cone of confusion will therefore mainly be referred to as front/back-confusion.

As mentioned, we also have certain abilities in localizing sounds with only one ear. The reason for this is that we have many different neurons responding to IID’s (lateral superior olive neurons, or LSO neurons), and ITD’s (medial superior olive neurons, or MSO neurons), and these do not only respond to differences between-, but also within the left and right ear ends (Heeger 2006). Figure 13 shows an example of the responses (or firing rates) an MSO neuron to detect ITD’s at either side of the head.

Interaural phase difference (IPD) is another cue related to the distances from the sound object to each ear, causing phase differences between the ears. For sine waves, a time difference equals a phase difference, and for low-frequency tones, the IPD provides an effective and unambiguous cue for the localization of the sound object. For high-frequency tones, however, the difference
can only be perceived for sounds up to 1500 Hz, as the distance differences between the ears for frequencies above this limit may lead to a phase difference equalling more than one period, causing ambiguous cues\textsuperscript{21}. During Sound Hunter’s game level construction, pure sine waves with frequencies above this limit were used (see the section on Planning Sound Hunter’s game levels), thus eliminating the possibility to use IPD’s/ITD’s as cues for navigational aid.

2.3.2.3 Spectral differences
Prior to reaching the listener’s eardrums, a sound it is affected by the head, pinnae (outer ears), and to some extent upper body shape. The sound is absorbed, diffracted, and transmitted around and through the body. Exactly how much the sound is affected depends on the sound source’s frequency and direction (Brieger & Göthner 2011). In other words, the head and pinnae form a direction-based filter, and transfer function for this filter is usually referred to as the HRTF (Head-Related Transfer Function).

In order to localize a sound object, humans interpret how sounds are affected by HRTFs, by comparing the above-mentioned differences between the ears, as well as timbre differences in the sound caused by the HRTFs. In order to be able to make timbre judgements, the listener must be somewhat familiar with how the sound object usually sounds (e.g. listening to a pre-recorded drummer in stereo headphones and then listening to the drummer in real-life) (Moore 2008). As HRTFs are based on the shapes of the listener’s head and pinnae, they may differ between listeners (i.e. using another person’s HRTF in order to localize sounds). This will be discussed in further detail in the spatial sound reproduction section below.

2.3.2.4 Further cues related to direction judgements
In addition to the cues having been mentioned, there are two further cues that aid sound localization, head movements and visual cues.

2.3.2.4.1 Head movement
Head movement is extremely important for auditory spatial perception, and can be seen as the main tool for eliminating the cone of confusion. If the location of a sound object is unclear while the head is in a certain position (i.e. several different locations are plausible, for example due to equal ITD’s), head movements will lead to the creation of new cues, thus making it possible for us to exclude the ambiguous cues and locate the sound object. The same result may also be achieved by moving the sound object as opposed to the head (Moore 2008)\textsuperscript{22}.

2.3.2.4.2 Visual cues
Visual cues are very effective due to the fact that if a sound is heard but not seen, the listener immediately knows that the sound is not located within the visual field, whereas if the sound source is both heard and located within the visual field, audio-visual cues are created, combining the strength of both auditory and visual cues\textsuperscript{23}. When it comes to visually impaired people, researchers have been arguing whether or not the auditory system’s ability to provide spatial presence is strengthened or not in the absence of visual information. Two main models have been proposed; the deficit model and the compensation model, both of which have received roughly equal support (D. Easton et al. 1998). The deficit model argues that visual information is encoded together in auditory information, and therefore acts as a reference system. Thus, if a person has no visual information available (i.e. no reference system), they will also suffer from impaired spatial hearing. The compensation model on the other hand, argues that in the absence of visual information, non-visual areas of perception in the brain may become more highly developed among visually impaired people compared to sighted people (D. Easton et al. 1998), (Rauschecker 1995). The compensation model can therefore be seen as an

argument that it in fact is possible to train auditory spatial perception, but today it is still rather unclear whether this may lead to permanent improvements also for sighted people.

### 2.3.3 Distance judgements

The human ability to make distance judgements when using our hearing is limited, and a 20% error margin is common. Also, we tend to exaggerate distances for sounds located close to us, and minimize distances for sounds located further away (Moore 2008).

When judging the distance to a sound object, the following cues are used:

- **Sound pressure level**: Sounds weaken as they spread, and the distance to familiar sounds can be determined by judging how strong they sound, especially in comparison to other sound objects and their reflections (i.e. reverberation, echoes). Therefore, the relationship between direct sounds, reflections, and resonance becomes an important factor when making distance judgements.
- **Changes in intensity (power)** are noticeable when moving towards or from a sound object, where less changes equals longer distances.
- **The sound’s spectrum/timbre**: The distance to familiar sound objects can also be judged by noticing the change in timbre, caused when the air absorbs the higher frequency components in the sounds.

For audio-visual computer games, it is generally not desirable to create realistic distance signals for sounds (Bogren Eriksson 2010). There are several reasons for this, for example that the overall sound level would become too high, and that certain sounds are more important than other sounds and therefore given more presence, even at further distances (Bogren Eriksson 2010).

The same applies for audio-only games, where additional sounds are added to aid navigation, as opposed to building a functional navigation relying on more realistic auditory cues, both of which have to be balanced correctly. This is one of the most important aspects behind the development of the framework for Sound Hunter, and will be discussed in greater detail in the section on my Proposed framework: Rethinking the design of 3D-audio games.

#### 2.3.3.1 The “Precedence effect”

Under normal conditions, the sounds we hear from sound objects are massively complex combinations of direct sounds, direct- and distant reflections from various objects in the environment, as well as resonance changes in all of the sounds in the environment. Despite this, humans still have the ability to determine the location of the sound objects, and the reason for this is the so-called precedence effect, which states that humans localize the sound object in the direction from where it was first heard (Moore 2008).

### 2.3.4 Other influences on our directional hearing

#### 2.3.4.1 Masking

Everyday, we experience sounds becoming inaudible due to the presence of other sounds, and this is referred to as masking. Masking most commonly occurs without us thinking about it. For example, if we stand in a room and clap our hands, the sound from the main clap will be the most noticeable sound. This is partly due to the above-mentioned precedence effect (i.e. the sound from the clap reaches our ears before the reverberation), but also because the clap masks other sounds.
the early echoes due to its higher intensity. This due to the two thresholds of masking (Moore 2008)\textsuperscript{28}:

**Unmasked threshold.** The quietest level at which the signal is perceived without masking.

**Masked threshold.** The quietest level at which the signal is perceived when combined with the masking signal.

Therefore, if the clap is recorded, reversed and then played back, we will instead perceive all the echoes followed by the clap, as the reversed signal never goes over the masked threshold. Masking may, however, also be used intentionally to avoid the perception of other sounds (e.g. turning up the radio in the car to avoid the sound of the engine). It could therefore also be used in audio game design to make certain sounds inaudible.

### 2.3.4.2 The “Doppler effect”

The Doppler effect occurs whenever there is a sound object moving with respect to a listener, thus altering the sound (e.g. when driving by, an ambulance will emit a higher pitch while approaching the listener, an equal pitch whilst at the listener, and a lower pitch while resending from the listener). The reason for this is that each wave reaches the listener faster when approaching (thus shortening the wavelength and making the pitch higher), and slower when resending (making the wavelengths longer and the pitch lower).

### 2.3.4.3 Diffractions

As mentioned, the head and pinnae cause interferences and diffractions in sounds, aiding our auditory localization. However, interferences and diffractions may also have other effects on our ability to localize a sound object regardless of the shape of our head and pinnae. For example, if a large physical object (such as a wall) is present in the environment, it bends the lower frequencies, while cutting off the higher frequencies, affecting our ability to localize the sound as low frequencies are over-represented compared to high frequencies, see Figure 14\textsuperscript{29}.

![Figure 14: Diffractions affect high frequencies more than low frequencies (source: http://hyperphysics.phy-astr.gsu.edu/hbase/sound/diffrac.html)](http://hyperphysics.phy-astr.gsu.edu/hbase/sound/diffrac.html)

### 2.3.4.4 Interference

Interference occurs when sound waves in two different sounds (either coming from different sound objects, or from reflections by the same sound object) have equal frequency components.

---

\textsuperscript{28} HyperPhysics, Diffraction of Sound, http://hyperphysics.phy-astr.gsu.edu/hbase/sound/diffrac.html

\textsuperscript{29} HyperPhysics, Diffraction of Sound, http://hyperphysics.phy-astr.gsu.edu/hbase/sound/diffrac.html
If their amplitudes add, *constructive* interference will occur, making the sound louder, whereas if their amplitudes are out of phase, *destructive* interference will occur, making the sounds quieter, or possibly eliminating the sounds completely, see Figure 15. When walking around in an echoic room where a sound is being played, it is therefore possible to determine the room’s *sweet spots* (balanced interference), *live spots* (constructive interference), and *dead spots* (highly destructive interference).

![Diagram of constructive and destructive interference](http://hyperphysics.phy-astr.gsu.edu/hbase/sound/interf.html)

Even though the auditory environments in audio games may be calculated using advanced models of real environments, both diffractions and interferences may still severely affect the perception of the auditory environment, regardless of the accuracy of the calculations or models (after all, these effects are natural, and occur in real life). Therefore, it becomes important to realize these effects when attempting to build audio-only games, as they might not always be desirable (see the section on Audio games).

---

30 HyperPhysics, Interference of Sound, http://hyperphysics.phy-astr.gsu.edu/hbase/sound/interf.html
2.4 Human vision

Even though this thesis mainly focuses on how the development of audio games may be improved by focusing more attention to human hearing and the processing of auditory information, I would still like to add a short section on human vision, where the focus is on various forms of visual impairment.

Visually impaired people are Sound Hunter’s main intended user group, and people with visual impairments have also been a very important part of the game’s development process from beginning to end. Working together with visually impaired people offers exceptionally valuable insights on how audio games ought to be be created in order to fulfil the usability criteria for both sighted and visually impaired users, and is something that I highly recommend to anybody attempting to create an audio game.

2.4.1 Visual impairment

Visual impairment is the condition of vision loss where there is a loss of visual functions at the organ level, and is defined in several stages, where all stages are connected with the measurement of visual acuity, or the spatial resolution of the visual processing system (Kettlewell 2002). Visual acuity is measured by letting a person identify characters on a chart from a distance of 6 meters and comparing the results with the degree of clarity that a person with normal vision would have (i.e. a person with normal vision has the visual acuity of 6/6 = 1, but having a visual acuity of 0.8 or over is usually also considered normal)\(^{31}\). Usually, the vision loss can be corrected using lenses, but if the visual acuity is 6/60 (0.1) or worse, this is referred to as legal blindness (i.e. the vision loss at which the person is legally defined as being blind). The normally used definitions of vision loss are (Kettlewell 2002):

**Corrective lenses**

- **Mild vision loss** – Visual acuity < 0.8
- **Moderate vision loss** – Visual acuity < 0.3

**Legal blindness**

- **Severe vision loss** – Visual acuity < 0.125
- **Profound vision loss** – Visual acuity < 0.05
- **Near total vision loss** – Near blindness
  - Visual acuity < 0.02
- **Total vision loss** – Total blindness, or “no light perception” (NLP)

Furthermore, *functional vision* is a term relating to the person’s ability to use vision in activities of daily living (ADL), which in some cases may downgrade moderate vision loss to legal blindness.

All of the visually impaired participants in the development of Sound Hunter had legal blindness.

2.5 Spatial sound reproduction

Spatial sound reproduction is the process of using speakers or earphones in a way such that the listener perceives the sound as coming from a certain direction and distance. There are many ways of reproducing sounds as being spatial. The most basic method, *stereo*, either through speakers or headphones, allows the control of the sound from left to right. More advanced speaker settings, such as *surround sound* (Murphy 2011), where the first number is the amount of speakers, and the second number is the subwoofer, introduce more azimuth angles. A more extreme (and unusual) speaker layout is that of *ambisonics* (Murphy 2011), where eight or more separately channelled speakers are positioned in a circle around the listener. For more details on how various speaker layouts may be used to create auditory spatial effects, I refer to the online book *An Introduction to Spatial Sound* by Damian Murphy.

In this report, the focus will be on spatial sound reproduction through headphones, mainly stereo and HRTF, whereas other speaker layout methods will be used as more of a reference in the methods section to indicate what types of spatial sound reproduction techniques the participants are used to or have experienced.

2.5.1 “Out-of-head experience”

When listening to sounds through stereo earphones, it is common to experience the sounds as coming from positions inside the head as opposed to various positions in space. This is referred to as *lateralization*, and is usually measured by letting the listener make judgements of the sound’s position on a straight line through the head between each ear. Lateralization differs from *localization*, where the listener perceives the sound as being in an actual position in space (Murphy 2011).

The ability to recreate an “out-of-head experience” (i.e. the sound is perceived as coming from a position outside of the head) is therefore an important quality measurement when studying auditory spatial techniques based on headphones.

2.5.2 Binaural sound and Head Related Transfer Function (HRTF)

2.5.2.1 Binaural sound

Binaural sound reproduction is based on recreating the navigational cues naturally used by humans when localizing sounds. As discussed in the section on directional hearing in humans, the sounds we hear are affected, or filtered, by the shapes of our head and pinnae. By placing microphones at the outer ear canals of either a human- or human-shaped upper body model, and then playing back the recording at the outer ear canals, the listener will receive all the directional information related to the filtering caused by the head and pinnae. The technique therefore only works when using earphones, headphones, or two extremely directed speakers. In other words, the closer to the outer ear canal the better, such as with in-ear earphones, not only because of the placement, but also because this helps avoid the signal from one ear entering the other ear.

As discussed in the section on directional hearing in humans, the head and pinnae cause three main differences between each ear: level differences, time differences, and timbre differences. The way these differences depend on the size and shape of the listener’s head and pinnae is very complex, and even though binaural recordings for one person are relatively simple to make and also give accurate results when recording from the same person’s head (making it an *individualized* HRTF), it is difficult to say how well this recording will work for other people (making it a *generalized* HRTF, also referred to as a *generic* HRTF or a *foreign* HRTF).

---

Studies have generally shown that there is a significant difference between using one’s own individualized HRTF compared to a generalized HRTF, but also that even without individualized HRTFs, people are reasonably accurate in localizing sounds when using generalized HRTFs, although not as accurate in determining front from back (Moore 2008)\textsuperscript{34}, (Wenzel et al. 1993)\textsuperscript{35}. Furthermore, studies have shown that a person can adapt to another person’s HRTF, indicating that we can change our spatial perception to that of another person and then train it until we perceive it as if it were our own (Begault, Wenzel & Anderson 2001). For example, Paul M. Hofman, Jos G.A. Van Riswick and A. John Van Opstal tested how long it would take for a person to adapt to a foreign HRTF during interactive daily experiences, by modifying participants’ pinnae with molds (i.e. creating a generalized HRTF). During a time period of six weeks, the participants’ accuracy in spatial perception was monitored, and compared to the baseline measurement consisting of a control group without mold implants. The results showed that azimuth localization was almost unaffected by the “new ears” immediately after the molds were inserted, whereas an accurate elevation localization could not be achieved until several weeks had passed (Hofman, Van Riswick & Van Opstal 1998). These results indicate that utilizing the elevation for HRTF-based audio-only games might not be possible, simply because it takes too long to adjust to the elevation of a generalized HRTF. However, as no studies have been conducted relating to the adaptation time, or learning curve, for azimuth localization during interactive gaming experiences, this was one area that I wanted to investigate when letting participants play Sound Hunter. Also, as Sound Hunter was developed using a generalized HRTF, all participants were tested for front-back confusion prior to playing the game, by using a graphical interface and simultaneous listening. However, Sound Hunter was not based on binaural sound recordings, but on synthesized HRTF filtering, being explained below.

2.5.2.2 Synthesized HRTF filtering

When it comes to computer games, it might not be possible to use binaural recordings, especially when attempting to create synthesized auditory environments changing realistically according to the player’s movements. Instead, generalized synthesized HRTFs are used, which are created by recording impulse responses from various positions around a human- or human-shaped upper body model, which are then used to filter any sound with the corresponding transfer function (see sections 2.1.6 Impulse responses and 2.1.7 Convolution). In other words, when wanting to position sound recordings or synthesized sounds (always in mono), they pass a filter corresponding to the HRTFs for the position where they are going to be placed (see Figure 16).

For example, when intending to place the sound of a talking baby at an azimuth angle of 20 degrees and an elevation of 40 degrees, the impulse responses (one for each ear) are needed for those angles. If no impulse responses have been recorded for exactly those angles, either the closest impulse responses, or averages of the closest impulse responses are used, where averaging is most commonly performed through linear interpolation (averaging the two closest directional impulse responses). Through convolution, the sound of the talking baby is then passed through the transfer functions for the right and left ear, including the corresponding level-, time-, and timbre differences. There are also more advanced ways of interpolating, such as the inverse distance weighting method, where several close HRTF’s are weighted by the inverse of the distance to the (to be obtained) HRTF (i.e. it accounts not only for the direction, but also for the distance, or even room acoustics, when interpolating) (Ajdler et al. 2005). In this thesis, linear interpolation was used, and therefore inverse distance weighting will not be covered in further detail.

---

\textsuperscript{34} Moore, B.C.J. (2008). \textit{An Introduction to the Psychology of Hearing}, p. 252.

2.5.2.3 Limitations

Apart from the already-mentioned limitation of poor perception of elevation for generalized HRTFs and the requirement of headphones or earphones in order to properly perceive binaural sounds or HRTF filtering, both methods also rely on the listener holding the head straight and still (unless a head tracker is being used). The reason for this is that the auditory environment is created from recordings around a still head, meaning that the environment is locked in this position (i.e. if the listener hears a sound object in front of them and leans the head forward, this will result in the sound object being perceived as being below and in front of the feet, unless the head movement is somehow connected to the auditory environment).

In the case of Sound Hunter, these limitations were taken into account by excluding the elevation parameter (as adjusting to the elevation would take too long for the participants), and the head movements corresponding to turning the head to the left or right were connected to the iPhone’s accelerometer data (see the results section on Sound Hunter: The original game idea). By doing this, the auditory environment changes according to the navigational input. However, as the navigational input is based on leaning the iPhone instead of turning the head (i.e. virtual head movements are possible, but the player still has to hold their physical head still), observations were made of all the participants’ head movements, to see if they compensated by trying to lean or turn their head in the real world outside the game.

2.5.2.4 HRTF filtering in Sound Hunter: [earplug~]

In Sound Hunter, the Pure Data external object [earplug~] (see the method section Pure Data), created by Pei Xiang, David Camargo, and Miller Puckette, was used to position sounds through synthesized HRTF filtering (Xiang, Camargo & Puckette 2005). Even though [earplug~] uses linear interpolation (i.e. it calculates new HRTFs through a linear interpolation of the closest impulse responses), it works surprisingly well, even in comparison to the more advanced commercial options for 3D-rendering, such as Rapture3D Advanced, which makes use of five
different HRTF sets from three countries, also performing advanced spatial calculations\textsuperscript{36}, or other simulations using inverse distance weighting. Through linear interpolation, [earplug]\textsuperscript{~} uses 368 (or 722 if mirrored to each side of the head) different impulse responses on a spherical surface (elevation of -40 to 90 degrees and azimuth of 0 to 360 degrees) gathered from KEMAR (dummy head) data sets (Gardner & Martin 1995). These are a lot of impulse responses for the direction, considering the fact that both listening tests and error analyses have shown that HRTFs may be shortened to 128 impulse responses and still give satisfactory results (which may be the way other systems work in order to reduce large data sets) (Sontacchi et al. 2002).

2.5.2.4.1 Limitations with [earplug]

As [earplug]\textsuperscript{~} uses linear interpolation, it only simulates direction and not distance. The distance simulation was therefore programmed separately (see the results section Optimizing distance), as opposed to interpolating distance and direction, as explained above (see Synthesized HRTF filtering).

Apart from this, the only problem found with [earplug]\textsuperscript{~}, is that it may give rise to imposed noise, either at very high frequency ranges (> 15 kHz), or if the quality of the audio sample is too low (usually resulting in high frequency distortions). All of the participants were therefore tested for whether or not they noticed the imposed noise for frequencies above 15 kHz and of lower quality sound files, and if so whether or not they were disturbed by it.

\textsuperscript{36} Blue Ripple Sound, Rapture 3D Advanced, http://www.blueripplesound.com/products/rapture-3d-advanced
2.6 Audio games

2.6.1 Introduction
Between 1996 and 2006, around 400 audio games had been developed, which is a very small number compared to visual computer games. The development teams were also small, usually consisting of one to four persons (Archambault et al. 2006). However the development has been more substantial over previous years, and the role of researchers, game developers, as well as sound designers has become more important in order to find more pleasant audio rendering techniques, as well as new exciting methods and technologies to use in audio games (Graeme 2011), (Archambault et al. 2006). There has been a rapid development of audio chips and 3D sound engines for computer games, and today, users as well as developers pay more attention to the audio content in computer games (Friberg & Gärdefors 2004). Sound is an expressive narrative medium, and sonic landscapes, or "soundscapes", may very well be as immersive and engaging as powerful 3D-based graphical environments (Friberg & Gärdefors 2004).

For games using both visual and auditory content where the main focus lies on the audio, the so-called “rhythm action games”, such as Dance Dance Revolution\(^\text{37}\), have become very popular (Friberg & Gärdefors 2004). In these games, the player’s input is meant to coordinate with the rhythm of a soundtrack. There are also examples of more simple portable audio-visual games having become very popular, such as SongPop\(^\text{38}\), which is available for all smartphone platforms and also integrates the dimension of social media.

However, despite the growing focus on audio in computer games, the audio content of mainstream computer games is extremely underdeveloped in comparison to the visual content (Friberg & Gärdefors 2004). This also applies to games being completely based on audio, which are extremely rare, compared to audio-visual games. Almost all audio-only games on the market today are made for PC computers, where some of the popular titles are SuperDeekout\(^\text{39}\) and Terraformers\(^\text{40}\). Furthermore, as most of the popular audio-only games are created in the first-person perspective, it also seems important to examine whether or not it would actually be possible to navigate in a game environment solely by using 3D-audio. This could possibly help in creating more exciting audio games, where the player no longer needs to rely on their own imagination of spatial audio, but instead is able to actually perceive the audio as being spatial.

2.6.2 Audio games and the visually impaired
Even though audio games are developed more sophistically these days, there is nonetheless very restricted access to an important part of the youth culture for people being visually impaired, and it could be argued that including this user group is of great importance, as it will aid their participation in society (Archambault et al. 2006). Accessibility to software applications is another area in which the visually impaired have more difficulties than sighted users. However, recent developments of various frameworks intended to make software applications more accessible has greatly aided the visually impaired, examples of which are Microsoft Active Accessibility, VoiceOver\(^\text{41}\) for the iPhone, as well as similar frameworks for Mac and Linux desktop environments (Archambault et al. 2006). By focusing more attention to game audio, new possibilities of designing games for people with visual impairments have emerged. However, this also requires that the games be developed with regard to their abilities and needs (Friberg & Gärdefors 2004). Also, as the number of blind iPhone users today is over 100


Audio games

thousand and rapidly growing\textsuperscript{42}, it therefore seems as if there is a significant and highly unfulfilled market potential for audio-only games intended for smartphone users.

\subsection*{2.6.3 Audio-only games}

Audio-only games are similar to video-based games, with the exception that they are played and perceived through sound and acoustics only (Röber & Masuch n.d.). Audio games have many advantages making them interesting for gameplay experimentation. For example, they allow an increased degree of spatial freedom, as no screens are necessary. Furthermore, the computational complexity is usually lower, meaning that less hardware is necessary, making them suitable for portable devices and mobile gaming (Röber & Masuch n.d.). Due to the lacking graphical representation, they may also increase the level of immersion, as the listener has to rely more on their own imagination, similar to when reading books (Röber & Masuch n.d.). Furthermore, games may also be used as a means to train the various senses and abilities in a person. For example, handicapped children may benefit largely both from using music (Hansen & Dravins 2012) and computer games in order to aid their psychomotor and cognitive development (Hildén & Svensson 2002), (Archambault et al. 2006). When it comes to audio games, these games can be used to train a person’s hearing and teach the player how to focus more on what they hear (Hansen, Li & Wang 2012).

\subsection*{2.6.4 Building audio-only games}

When building an audio-only game, it is important not only to use self-explanatory sounds, but also to establish agreements early on in the game so that all the necessary information is conveyed correctly to the player (Friberg & Gärdefors 2004). These agreements should build upon metaphors and associative patterns in order to make it easier for the player to get a sense of what information is important in the game (e.g. the difficulty of the challenge ahead, the current success rate, or the scores awarded when completing a game task). It is also important to emphasize the difference between various types of auditory information. For graphics, variations in colours, borders, buttons and other types of design principles are used to label and categorize different types of information. For auditory information, different models have been proposed, where one example is the SITREC categorization system suggested for audio game interfaces, including (Friberg & Gärdefors 2004):

\begin{itemize}
  \item \textit{Avatar sounds} – Represent avatar activity, such as footstep sounds
  \item \textit{Object sounds} – Indicate the presence of objects, which can be brief, recurring, long, or continuous sounds depending on the presentation strategy
  \item \textit{Character sounds} – Sounds generated by AI
  \item \textit{Ornamental sounds} – Not necessary for conveying gameplay information, such as ambient background music, and
  \item \textit{Instructions} – Usually speech recordings
\end{itemize}

However, it is also important to distinguish auditory information generated by player activity from the information generated by other sources in the game. This player feedback informs the player whether or not the action was registered by the system (Friberg & Gärdefors 2004). This has led to the development of three well-established design methods for auditory interfaces:

\textbf{Auditory icons} are recognizable sounds, such as voices or confirmatory sounds (Gaver 1993), (Helander, Landauer & Prabhu (eds.) 1997)

\footnote{\textsuperscript{42} Texas School for the Blind and Visually Impaired. EyeNote research (BEP). \url{http://www.tsbvi.edu/component/content/article/182-fall-2011/3617-eyenote, 2013.}}
**Earcons** are short musical phrases (Brewster 1998), which similarly to auditory icons are associated with various types of information to inform the player of their actions (Friberg & Gärdefors 2004).

**Sonification** can be seen as sub-part of the auditory icon, and is the process of mapping abstract data to non-speech sound (Röber & Masuch n.d.).

### 2.6.5 Meta-level communication and sound characteristics

Additionally, it is possible for game sounds to have several layers of information, often when wanting to enhance the complexity or function of the sound. This additional information may be communicated on a *meta-level*, for example when wanting to communicate the level of activity in the sound of a motor (Friberg & Gärdefors 2004). Meta-level communication was very important for the construction of the synthesized game levels in Sound Hunter (see the section on *Planning Sound Hunter’s game levels*).

When considering the sound characteristics, it is also important that each sound is intelligible and distinguishable. Often, the sounds are also accompanied by a musical context, the latter of which seldom is emphasized, but rather added as a backdrop in order to create a scene and set the mood (Friberg & Gärdefors 2004), (Graeme 2011). Another aspect of sound characteristics is that of looping sounds, which may reduce the level of realism in the sound, but is still often desirable in audio games, as it gives the player an overview of the game space (Friberg & Gärdefors 2004).

### 2.7 3D-based audio-only games

In many audio-based action games, the success of the player depends on interaction based on precise timing (Archambault et al. 2006). Examples of audio-based action games including 3D-audio are most commonly found in the first-person shooter genre, for example *Shades of Doom*[^43], *Terraformers*[^40], and *Demor*[^44], all of which are made for the PC. Another popular audio game category is that of adventure games, or exploration games, where three key features are combined; an interesting scenario, the exploration of new worlds, as well as activities of riddle solving (Archambault et al. 2006). Examples of audio-based exploration games including 3D-audio are *Blindside*[^45] and *Escape The House: A 3D Sound Experience*[^46], both of which are available for iOS users.

#### 2.7.1 Current problems with 3D-based audio-only games

Röber and Masuch attempted to prototype various game ideas for audio-based gaming by using individualized HRTF’s, head tracking, and a joystick or keyboard for player movement (Röber & Masuch n.d.). They created three action-based games (*The Frogger Game Remake*, *Mozquitos*, and *MatrixShot*) and one exploration game (*The hidden Secret*). Röber and Masuch pointed out several future improvements, such as extending their framework with more advanced sonification and interaction techniques, as well as developing a truly mobile solution, allowing the player not to be bound by webcams, head tracking devices and other stationary equipment. These limitations are also found in the more mainstream audio game *Demor*[^7] (Röber & Masuch n.d.).

In his Master’s degree project, Graeme developed a 3D-audio game called *Blind Fear*, featuring advanced auditory environments including appropriate reverberation and occlusion effects (Graeme 2011). However, similar to the previous studies having been mentioned, Graeme found

that the more effort he put into creating auditory environments being as advanced and rich as possible, the more difficult it became to actually manage to navigate in the auditory world. During various parts of the development process, Graeme managed to play through the entire game from beginning to end using only audio, but only after simplifying it by sacrificing additional sounds, thus compromising the aesthetical effort he had put into the game (Graeme 2011). This example not only shows the difficulties that may arise when trying to balance functionality and aesthetics in an audio game, but also the need of including the intended users as early on as possible in the game’s development process to minimize compromises between usability and aesthetics.

One of the more promising games is *Blindside*, requiring only a smartphone and headphones and where only a few sounds are focused on at a time, making the 3D-audio easier to perceive. However, there are still problems with this game. For example, it is quite static and slow, just like many other audio games (e.g. press forward button to move in the game at a pre-determined speed). *Blindside* also relies on the iPhone’s gyroscope (where the player needs to stand up and spin around), which on the one hand gives more spatial freedom, but on the other hand makes it difficult to play when sitting up or lying down.45

For the more mainstream 3D-audio-only games not relying on stationary equipment (e.g. *Shades of Doom*46, *Terraformers*40, or *Escape The House: A 3D Sound Experience*46), the actual 3D-effects are difficult to perceive. The games give the impression that they have been developed as if they were intended for stereo usage, with the exception of various additional binaural sounds introduced to the mix. As there are many tools and programming libraries available for creating 3D-audio environments, such as FMOD’s head related transfer function (Graeme 2011), and it can therefore be expected that developers and sound designers rather decide to enhance their stereo environments with various binaural sounds, than to rely merely on stereo usage. However, when attempting to build navigational 3D-audio games, it becomes extra important not only to add filtered 3D-audio, but to also to understand the filtered sound’s relationship to other sounds in the game, in order to design the auditory environments such that they are perceived and interact appropriately to the user’s input. In the following section, I present a design framework for how navigational 3D-audio games might be built to accomplish this, and the framework also acted as a foundation when developing Sound Hunter.
3 Proposed Framework

In this section, the current problems with 3D-based audio-only games are summarized, and the proposed design framework for navigational 3D audio games is presented.

3.1 Rethinking The Design of Navigational 3D-Based Audio-Only Games

All of the above mentioned audio-only games based on 3D-audio claim to present the player with exciting 3-dimensional sound-environments, either through surround sound systems, or by 3D-audio through headphones. However, the focus is usually not on the player’s ability to perceive the 3D-audio in order to use those cues as the main tool for navigation, but rather on enhancing the perceived quality of the stereo environment in order to make it richer and more life-like. For the games focusing more on the navigational purposes of the 3D-audio, the problem instead seems to be their reliance on head tracking devices or other stationary equipment (Röber & Masuch n.d.), making them highly inappropriate for relaxed gaming (e.g. lying down as opposed to walking around), or mobile gaming with a smartphone and earphones.

In order to truly make use of 3D-audio for the purpose of navigation, it becomes very important not only to create the 3D-audio environments themselves, but to also understand our natural abilities and limitations when it comes to perceiving spatial sounds as human beings. For example, when filtering a sound using an HRTF filter in order to place the sound in a certain position in a 3D space (e.g. in front of the listener and slightly to the right), the sound will not automatically be perceived as being in front and slightly to the right of the listener without involving movement, either by moving the player’s (head) position, or the sound source’s position (see the section on directional hearing in humans). As mentioned in the section on Directional hearing in humans, the cone of confusion arises when sounds in different positions in a vertical circle around either side of the listener’s head have equal inter-aural time differences (ITD’s), and the confusion is eliminated when moving the head (i.e. causing the sound’s relative position to the listener to change, thus altering the ITD’s) (Moore 2008).

Furthermore, additional stereo sounds being used in audio games may disturb the player’s ability to perceive the HRTF filtering. This is because our ability to perceive a sound as being spatial depends on the above mentioned ITD’s, as well as inter-aural intensity differences (IID), and further qualities of the sound, such as spectral differences, or room qualities simulated by reverberation, all of which may be masked, interfered, or in other ways become inaudible in either ear, causing severe inability to localize the sound object (Graeme 2011), (Moore 2008). As many audio-only games using 3D-audio can have over 30 different tracks playing simultaneously (Helander, Landauer & Prabhu (eds.) 1997), where some are in stereo (e.g. game music, instructions, auditory icons or earcons, usually at a relatively high volume), and some may be HRTF filtered, it is therefore not surprising that the 3D-effect becomes difficult to perceive.

There is usually a trade-off whether to use more sounds to enrich the auditory environment, or to use fewer sounds to aid navigation (Friberg & Gärdefors 2004), (Graeme 2011). Most common, however, is to add more sounds, such as auditory icons and earcons, in order to aid navigation through the already over-complex auditory environments (Röber & Masuch n.d.), (Friberg & Gärdefors 2004). The framework for Sound Hunter was therefore developed in a way ensuring that the 3D-audio navigation was the most important aspect of the game, with further sounds being added only later on in the development process.

Finally, there exist, to the best of my knowledge, no examples of audio-only games where it is clearly stated that the game was developed in an iterative process together with visually impaired people (or other intended user groups for that matter). As modern technology-oriented
development processes based on Human Computer Interaction (HCI) strongly suggest that the intended users should be part of even the earliest Lo-Fi prototypes (e.g. paper prototypes, workshops) (Rettig 1994), I therefore suggest that this is one of the general key aspects of importance in developing a highly functional and entertaining audio game, satisfying its intended user group.

The proposed framework can be viewed in Figure 17. Especially important are the recurring iterations of in-game navigation being used to update the sound design aspects. These iterations and evaluations should be made with the target user group.

Figure 17: Proposed framework for developing navigational audio-only games based on 3D-audio.
4 Method

4.1 Why use HCI methods?

4.1.1 Human Computer Interaction

Before I begin explaining the methodology used to create Sound Hunter, I would like to give a more thorough overview of the already introduced concept of Human Computer Interaction (HCI), as well as which parts of it are the most relevant for this thesis. By definition, HCI can be seen as:

“Discipline concerned with the design, evaluation and implementation of interactive computing systems for human use and with the study of major phenomena surrounding them” – (Gulliksen & Göransson 2009).47

In other words, it is a multidisciplinary study, involving not only the design of the computation devices themselves, but also the (usually unforeseen) phenomena surrounding them, such as how users will interact with the system, or environmental aspects affecting (or being affected by) the system. Ignoring these surrounding factors in a system’s design is likely to lead to a poor product and higher development costs, and HCI can therefore be seen as a method used to ensure that the system or technology is developed such that it fulfils its requirements without any second-guessing.

4.1.2 Mixed methods design

While qualitative research has the pros of getting an understanding of the underlying reasons, motivations, opinions, or insights into why the problems exist, it has the cons of small sample sizes, where results cannot be generalized over a larger population of interest, which is answered by quantitative research (Abeyasekera 2000). Quantitative research on the other hand, has the pros of quantifying data and generalizing the results of a sample to a larger population of interest, while the cons are connected to the inability to understand complex human behaviour, behavioural changes, and why certain findings were found, which is answered by qualitative research (Mitchell, Max & Joanne 2003).49

In development processes, it has therefore become increasingly common to use mixed methods as a research method, and as the name suggests, it can be seen as mixing qualitative and quantitative methods. The reason why this is usually preferable, is (as mentioned above) that qualitative and quantitative research methods give different results, which in certain cases makes the combination of both methods stronger than choosing to focus on only one of them. However, designing mixed methods research is usually more challenging than expected. This is because most development projects are constrained by a time frame, meaning that there is no time to deeply examine both methods, and instead the most important parts of the qualitative and quantitative methods in connection to the development process have to be chosen. Furthermore, it might be impossible to know beforehand what the most important parts are. Depending on the research question, one method is therefore usually chosen first and the second added later on in the process to strengthen the results of the first method.

In the case of Sound Hunter, I knew beforehand that the sample sizes would be small, as the user group is very narrow (visually impaired people also being interested in audio games). Therefore, I chose to use qualitative methods first, followed by quantitative data analyses of certain aspects connected to the usability tests.

---

4.1.3 Iterative development

Iteration means repetition, and the foundation of an iterative development is the insight that something cannot be made perfect on the first go (Gulliksen & Göransson 2009). Initially, it is impossible to know exactly what characterizes the problems, or what the demands are. Instead, the philosophy is repetition and refinement until the iteration is reached where the system is effective, efficient, usability-tested, and provides satisfaction to its intended users. When developing technology intended for somebody else, it is therefore always best to develop it in several iterations together with the intended users.

In the case of Sound Hunter, I had no familiarity with audio-only games, and I had never even met a blind person before. Therefore, reaching out to the user group and working with them iteratively came naturally, and was an extremely important part of the game and framework’s development, as it gave me additional information from the users that I would never have got otherwise.

4.1.4 Pure Data

Pure Data (Pd) is a visual open source data flow programming language, where the developed software patches are represented graphically. Algorithmic functions are called objects (usually represented in square brackets, e.g. [object_name]), which are placed on a canvas screen and are tied together with cords representing the data flow from one object to the next. An object may perform anything from low-level mathematical operations, to complex high-level audio or video functions (e.g. reverberation, FFT transformations, or video coding), and are either built-in Pd objects, abstractions (reusable patches built inside Pd itself), or externals (objects developed in another programming language).

When developing Sound Hunter, I made use of all of these types of objects (e.g. the Pd external earplug~, low-level built-in mathematical objects, and my own-built abstractions connecting everything together). When explaining Sound Hunter’s post-development stages, I will mainly be covering the abstractions having been built (e.g. navigation, game levels, highscore calculations, auditory icons).

4.1.5 Lo-Fi and Hi-Fi prototyping

The concepts of Low Fidelity (Lo-Fi) and High Fidelity (Hi-Fi) prototyping are usually connected to iterative development processes. A Lo-Fi prototype essentially means creating an extremely easy prototype, which can be made even by pen and paper by the users themselves (Rettig 1994). Hi-Fi prototypes on the other hand, are more highly developed, usually created towards the end of the development process, and can either be vertical (focusing on a small part of the total system in full detail, i.e. Hi-Fi), or horizontal (giving a general overview of the entire system, but not including all the functionality, i.e. either Lo-Fi or Hi-Fi).

During Sound Hunter’s iterative development process, a horizontal Lo-Fi prototype was first designed, followed by developing a vertical Hi-Fi prototype to test the navigation, and finally a horizontal Hi-Fi prototype to test the entire game.

4.1.6 Reliability and validity

When choosing methods for collecting and analysing scientific data, it is important to consider how reliable the information is, as well as if it can be seen as valid (Bell 2009). Reliability essentially means that the methods and instruments chosen in the experiment should be as consistent and trustworthy as possible, meaning that the experiment gives the same results when being performed again. In other words, the reliability is the trustworthiness of the methods and instruments used in the experiment. Validity is a measurement of the extent to which an experiment actually answers the intended questions, and is a more complex measurement than

---


Method

reliability, as the perceived validity of an experiment may vary from person to person (Bell 2009).

In the case of Sound Hunter, the methods chosen for the development, testing, and evaluation of the game are well-established HCI methods (e.g. focus groups, usability tests, interviews). The reliability therefore mainly depended on the implementation of the game. However, as Sound Hunter was developed with open source programming (i.e. there is no copyright protection of Sound Hunter’s content), the tests having been performed could easily be replicated.

Furthermore, as the participants were chosen from a very narrow user group, as well as the fact that the experiments and programming are explained in great detail, the reliability was considered high. Regarding validity, the consistency of the observations and the participants’ statements for the qualitative data analyses, the implementation of the quantitative data analyses (e.g. including significance values and confidence intervals), the combination of methods used throughout the process, as well as the fact that the whole iterative development process was founded on empirical studies with the resulting framework, supports that the results are valid.

4.1.7 Effectiveness, efficiency, and satisfaction

The usability of a technical prototype is typically measured in its effectiveness, efficiency, and satisfaction (Frøkjær, Hertzum & Hornbæk 2000). While efficiency is something that can be measured quantitatively (e.g. the extent to which cost, effort, or time is well-used in order to achieve a certain task), effectiveness is more vague and non-quantitative, generally concerned with achieving goal objectives. Efficiency is therefore commonly referred to as “doing things right”, whereas effectiveness is referred to as “doing the right things”52. User satisfaction can be measured either by quantitatively structuring qualitative data (e.g. through measuring and comparing answers on 7-point Likert scales), or by using qualitative content analyses (e.g. semi-structured or open questions).

Effectiveness: For Sound Hunter, the effectiveness concerned how well the developed prototypes, as well as the finished game, answered the various research questions in connection to the three main areas of research: 1) Developing a navigational HRTF-based audio-only game without relying on stationary equipment or physical movement, 2) If there is a need for audio-only smartphone games, and in this case how Sound Hunter should be developed in order to be as compatible as possible with the way in which blind people use their smartphones, and 3) How HRTF-based audio-only games may be built to train a person’s hearing in an entertaining way. The effectiveness can therefore be seen as the generalizability of the suggested development framework, the results from each of the iterations throughout the development process, as well as the summarizing conclusions discussing the broader scope and overall generalizability of the results for similar developments.

Efficiency: The efficiency was measured through a series of structured tasks during focus group sessions 2-3, as well as the usability tests, and mainly concerned various aspects of spatial perception and completion times (e.g. front/back-confusion, the time it took to capture a sound object, the time it took to complete the various game levels, the participants’ mean navigational learning curves, and the extent to which the participants improved in the game over time). In focus group sessions 2-3, the completion times were measured mainly to give indications of various improvements related to efficiency. In the usability tests, the completion times were measured quantitatively to find significant results and trends related to efficiency.

Satisfaction: User satisfaction was measured qualitatively through content analyses of the semi-structured focus group sessions 1-3, in the usability tests through observations and semi-structured post-game interviews, as well as the semi-structured final evaluation. The participants stated most of their satisfaction as answers to questions after navigating or playing the game, but also through their behaviour and statements being observed while the participants played Sound Hunter.

4.1.8 Overview of Sound Hunter's development process

Sound Hunter’s game development process consisted of four parts involving 10 participants. The first two parts were connected to game development in order to develop the prototype, whereas the following two parts were connected more to game testing and evaluation, in order to perfect the existing prototype as much as possible during the time frame of the thesis, as well as to find areas of future development of similar games, and suggestions for future research:

Game development

1) **Pre-study** in which other 3D-audio games were evaluated in order to create the framework and come up with an initial game idea

2) **Three focus group sessions** with two middle-aged visually impaired participants (i.e. the development team), where each session was iterative and followed by game post-developments

Game testing and evaluation

3) **Usability tests** with eight (new) players: two middle-aged visually impaired players, two younger visually impaired players, and four younger sighted players, and lastly a

4) **Final evaluation** of the game with one younger visually impaired participant (also part of the usability tests)

Because of great difficulties in acquiring visually impaired persons wanting to participate in the development of an audio game, the evaluation methods were mainly qualitative. However, some quantitative data-collection and analysis was also performed during the usability tests to determine the game level order by difficulty, as well as to find trends in how the players gaming abilities improved over time. Due to the large quantities of qualitative data, only the most relevant quotes in connection to important discoveries having been made are mentioned.
5 Results

The results section is split into two parts. The first part is connected to Sound Hunter’s development process, and the second part is connected to testing and evaluating Sound Hunter. First, the original game idea for Sound Hunter is presented, followed by the adjustments having been made to the game throughout the development process. Following this, suggestions for optimizing Sound Hunter and ideas for future development will be presented.

5.1 Part 1 – Focus groups: Developing Sound Hunter

5.1.1 Sound Hunter: The Original Game Idea
The original game idea for Sound Hunter was relatively simple. My intention was to create a single-player action-based game in which a sound is placed randomly in a 3D-space, utilizing the azimuth angle, where the player’s objective would be to capture the sound as quickly as possible by using 3D-audio as the only navigational aid. The player’s movement was meant to be controlled by the iPhone’s accelerometer data, where leaning the iPhone forward or backwards (the accelerometer’s y-data) would lead to the player moving forward or backwards in the game, and leaning the iPhone left or right (the accelerometer’s x-data) would correspond to the player’s head turning either to the left (making the sound objects spin to the right) or turning the head to the right (making the sound objects spin to the left), as in real life. Once the sound had been reached, a confirmatory sonification sound would be heard, followed by the next level beginning, where each level would become progressively more difficult. The game would be programmed in Pure Data\textsuperscript{53}, where the accelerometer data would be sent to Pure Data using the iPhone application TouchOSC\textsuperscript{54}. If there were enough time, the game would be made available as a smartphone application by using LibPd to incorporate the Pure Data functions into other code\textsuperscript{55}.

5.1.2 Focus group sessions
The focus group sessions were semi-structured (i.e. specific questions were discussed) and iteratively re-occurring throughout Sound Hunter’s development process. Focus group sessions two and three also included a series of structured tasks in order to find flaws in various parts of the game, and to measure efficiency (i.e. spatial perception among the participants, the quality of the HRTF filtering, in-game navigation, task completion time, and auditory icons). All of the focus groups will be presented and explained, followed by the qualitative data analysis, results, as well as the subsequent post-development of the game.

5.1.3 Participants
The participants were the same for all three focus groups (i.e. the game’s development team):

Participant 1 (P1): Male, 31 years old, born with total blindness, no documented hearing impairments. P1 had played the occasional audio-based game, but never in depth. He also

\textsuperscript{53} Pure data, http://pure-data.info. A planned development is to use libPd to have the application running on a mobile device.
\textsuperscript{54} Hexler.net. TouchOSC: A modular OSC and MIDI control surface for iPhone/iPod Touch/iPad, http://hexler.net/docs/touchosc
\textsuperscript{55} LibPd: About, http://libpd.cc/about/
owned and used an iPhone, and worked as a usability expert for utilities intended for blind people.

Participant 2 (P2): Male, 57 years old, had a form of progressive visually impairment, where the most severe vision losses appeared about 20 years earlier, no documented hearing impairments. P2 had played simple arcade-like games in his youth, but had no experience with audio-based games. He did not use a smartphone, but instead an older Nokia. He was also very interested in music, and played in a band together with P1.

Both P1 and P2 had, coincidentally, experienced binaural sounds through the album A Momentary Lapse of Reason by Pink Floyd, but had no further familiarities or interactive experiences with HRTF synthesis or binaural sounds.

The participant selection was conducted in Stockholm, Sweden through the company Funka Nu.

5.2 Focus group session 1

My intention for this focus group session was to present HRTF technology, its capabilities, and find out the participants’ opinions on its potential usage in audio games. In order not to bias the participants, an open discussion was held related to the types of games that might be created using synthesized HRTF filtering, after which the initial game idea of Sound Hunter was presented and discussed. This allowed the gathering of general data and data being specific to the initial game idea.

5.2.1 Qualitative data analysis

The qualitative data consisted of the responses from the participants, and was recorded using an OLYMPUS WS-450S Digital Voice Recorder. In order to find the most relevant participant quotes, the qualitative data was transcribed and categorized according to the semi-structured topics in connection to both general and specific game development areas. As paying attention to data structure at the time of collection greatly facilitates a proper data structure when analysing the data (Abeyasekera 2000), this was therefore performed as a form of content analysis, in order to make it easier to find hierarchies of importance, as well as to find various trends in the data. The semi-structured templates (in Swedish) for all of the qualitative methods used in this thesis can be found in the appendix section at the end of the thesis.

5.2.2 Results

After some open discussion, both participants suggested that basically anything using this technology that is fun, entertaining and functioning has the potential of becoming a great success. P2 also pointed out that more simple games could be fun, such as the old-school arcade games he had played in his youth, in which the player reaches higher levels under various constraints (e.g. time, enemies). When presenting the game idea for Sound Hunter, both participants regarded the navigational aspect of being able to locate and control sounds in a 3D-based sound-environment as exciting (e.g. being able to hear sounds coming towards, or appearing behind the player). An important conclusion from this focus group session was that there would be no need for the instructions or highscore points to be pre-recorded. Instead, this information can be shown as text on the iPhone, as it allows the user to control the VoiceOver reading speed as they prefer. Furthermore, we concluded that the menu would have three items (“Play”, “Instructions”, “Highscore”), that a Hi-Fi prototype of the navigation would be developed for the next session, and that 20 game levels would be enough for further tests, where these levels could be altered in difficulty in several ways, such as:

1. Varying the level of realism in the sound (realism as in its connection to a real-life scenario e.g. crying baby vs. growling tiger),
2. Inserting different amounts of temporal silence in the sound,
3. Varying the complexity in the sound (e.g. dynamic drum-set vs. sine wave),

36
4. Varying frequency (e.g. different ranges or by approaching our perceptual limitations), and
5. Using additional cues related to distance changes being communicated on a meta-level (e.g. varying the level of activity or frequency in a sound object depending on its distance).

5.2.3 Post-development

5.2.3.1 Creating the first Lo-Fi menu prototype
The first Lo-Fi prototype of what Sound Hunter’s menu might look like on the iPhone was vertical, and included three items: “Play”, “Instructions”, and “Highscore”, as shown in Figure 18.

![Sound Hunter Menu Prototype](image)

Figure 18: Sound Hunter’s first menu prototype

5.2.3.2 Developing the Hi-Fi navigation prototype
Sound Hunter’s foundation is having a properly functioning navigation, and there were several ways in which this could be accomplished. The idea was to only be able to capture sounds from the front or back within a certain radius range, where sounds would be spun by the accelerometer’s x data (left/right), and approach or move away by the accelerometer’s y data (front/back). The simplest way of doing this would be to only allow the player to approach a sound when having spun it such that it is positioned exactly in front or behind them. However, as I also wanted the sounds to be able to fly past the player in order to give a proper illusion of motion and movement in all directions, it became slightly more complicated. The final Hi-Fi prototype, [navigate_turn_and_walk_around] (see Figure 19), therefore included four continuously changing variables: x (the sound object’s movement left/right), y (the sound object’s movement forward/backwards), phi (degrees/direction), and r (radius/distance). Note that x and y here are not the same x and y as the iPhone’s accelerometer data, but are calculated from the iPhone’s accelerometer data in the abstractions [turn] (x) and [walk] (y). Two other important objects here are [poltocar] (Polar to Cartesian values), and [cartopol] (Cartesian to Polar values).
Method

[navigate_turn_and_walk_around]

When a sound is heard the listener turns the iPhone left/right to affect the azimuth, moving the sound object around the listener. Leaning the iPhone forward/backwards will make the listener walk forward/backwards, where the sound object can move past the listener on the left or on the right hand side if not within 2% of the total radius (or width), in which case the sound will be captured. This works in Pure Data according to the following five steps:

1. The sound object is placed by setting a value for the radius (distance) and the degrees (direction), which is done by the yellow button to the left.
2. Leaning the iPhone left/right affects the set value for the degrees (blue horizontal scale to the left), but the set radius (orange horizontal scale to the left) remains constant at the maximum distance of r=1, or 8 meters (i.e. the sound moves around the listener in a perfect circle). The values for the degrees and radius are then used as input in the object [poltocar], giving the new x and y values from spinning the sound (white vertical and horizontal scales to the left).
3. The new x value is used directly as input in [cartopol] (blue horizontal scale to the right), and the new y value is sent to [walk]. Leaning the iPhone forward/backwards will then affect the y position in [walk], which in turn sends the new y value to [cartopol] (orange vertical scale to the right), giving the new values for the radius and degrees from walking (white horizontal scales to the right).
4. The new radius value is sent directly as input to [poltocar], and the new degrees value is sent to [turn] (thus updating the old degrees value so the sound object spins from the updated direction, received after walking). Leaning the iPhone left/right then updates the degrees again, sending them from [turn] as input in [poltocar], giving the new x and y values (i.e. the same as step 2, but with the new direction and distance).
5. Steps 2-4 are then repeated until the sound object has been reached.

Figure 19: Overview of the [navigate_turn_and_walk_around] object.

---

38
5.2.3.3 Planning Sound Hunter’s game levels

The initial plan for the game levels in Sound Hunter was to divide them into sound files (or audio loops) being mixed in Ableton Live\textsuperscript{56}, and synthesized sounds (sounds being created from scratch in Pd). The reason for this, was that I wanted to see how these different groups of sounds may be used to alter the difficulty level in the game, as well as to find out which types of sounds would be the easiest to localize. Also I wanted to find out the differences between these groups of sounds when it came to the participants’ abilities to localize them, as well as their opinions on which sounds were the most fun in the game.

**Sound files**: Sound files consist of pre-recorded audio, and can therefore be any kind of sounds (e.g. voices, growling beasts, singing birds, or instruments). They are therefore very useful, both because they capture a wide range of different sounds, and also because they are most likely recognizable or imaginable by the player, as they may connect to various real-life events or experiences. However, even if a sound in itself may sounded real (e.g. a gun shot), the realness of the sounds in Sound Hunter was determined according to their realism as their connection to everyday scenarios or experiences, such as:

- **Realistic game scenarios** (e.g. walking towards a crying baby)
- **Semi-realistic game scenarios** (e.g. walking towards a growling tiger), or even
- **Non-realistic game scenarios** (e.g. walking towards a screaming dinosaur)

**Synthesized sounds**: Even though synthesized sounds are not captures of real audio, they can be created in such a way that they either simulate real sounds (e.g. creating a humming motor by adding reverberation to a square wave with low cut-off frequency, thus creating a more complex sound), or in a way such that they sound completely different from real sounds (i.e. they are audibly perceived as being synthesized, for example by playing a single sine wave). Even though sound files have a definite advantage over synthesized sounds when it comes to capturing realism or being perceived as sounding more perfected (i.e. of higher quality, even if in reality the sound quality may be the same), synthesized sounds still have other advantages over sound files that I found could be useful in Sound Hunter. For example, synthesized sounds:

- Allow complete control over every parameter in the sound (e.g. by changing the pitch, or adding/subtracting infinite amounts of extra layers/oscillators to the sound to make it more complex, for example creating the sound by adding one sine wave oscillator, one square wave oscillator, and one triangular wave oscillator, or adding envelopes)
- Can be connected to other variable or non-variable game parameters (e.g. by creating a clicking or pulsating sound and making the clicking frequency vary depending on how close the sound object is to the listener, in other words connecting the clicking frequency to the radius variable, thus communicating meta-level information related to distance other than loudness)
- May be effective for training a person’s hearing (e.g. by creating sounds moving towards our natural hearing limitations as humans, for example sounds moving down towards 20Hz, or sounds moving up towards 20kHz)

The initial idea was therefore to create ten sound file levels and ten synthesized levels, where the difficulty of each level would be determined according to five different parameters in the sound (easier vs. more difficult):

1. **Its connection to real-life scenarios (realness)**: Strong vs. loose connection

\textsuperscript{56} Ableton Live, main website https://www.ableton.com/en/live/
2. **The level of oscillation (temporal variability in the sound):** Big vs. small changes to the characteristics of the sound over time

3. **The amount of meta-level communication in the sound:** Complex vs. less complex changes to the sound depending on its distance (e.g. level of activity vs. frequency)

4. **The amount of silence in the sound:** Constant playback vs. playback every now and then

5. **Frequency:** Wide vs. narrow range, or audible vs. less audible

From these requirements, I created ten sound file levels, and nine synthesized levels (see Table 1 and 2). However, when creating these levels, I noticed that the sound quality of two of the sound file levels (level 4 and 8) might not be satisfactory. I also noticed that level 15 and 18 led to slight noise when being filtered. Furthermore, levels 18 and 19 were very effective when it came to being difficult, but I realized that these levels might lead to confusion, as the meta-level communication in these levels led to the tones becoming increasingly difficult to perceive the closer they came (as opposed to levels 15 and 16, where the meta-level communication led to the tones becoming increasingly easy to perceive the closer they came). I therefore wanted to test these levels in the second focus group, in order to find out whether or not they had to be altered in some way, or changed for new levels.

<table>
<thead>
<tr>
<th>Sound files</th>
<th>Realness</th>
<th>Oscillation</th>
<th>Silence</th>
<th>Frequency range</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Drum-set</td>
<td>High</td>
<td>High</td>
<td>None</td>
<td>Very wide</td>
<td>Slight comp. / Reverb</td>
</tr>
<tr>
<td>2. Spooky chains</td>
<td>Medium</td>
<td>High</td>
<td>None</td>
<td>Wide, peak at 8 kHz</td>
<td>Slight reverb/ chorus</td>
</tr>
<tr>
<td>3. Baby talking</td>
<td>High</td>
<td>High</td>
<td>≈ 1.5 seconds, split up</td>
<td>≈ 500-5 kHz</td>
<td>Dynamic EQ, slight reverb</td>
</tr>
<tr>
<td>4. Fly buzzing</td>
<td>Medium</td>
<td>Low</td>
<td>≈ 0.5 seconds</td>
<td>Medium, main content around 1-2 kHz</td>
<td>High-pass filter, slight reverb</td>
</tr>
<tr>
<td>5. Spooky water drops</td>
<td>Low</td>
<td>Medium</td>
<td>None</td>
<td>Wide (echo), drop peak at 900 Hz</td>
<td>Echo (between drops), slight chorus</td>
</tr>
<tr>
<td>6. Raven</td>
<td>High</td>
<td>Medium</td>
<td>≈ 2 seconds</td>
<td>Wide, main content around 800-4 kHz</td>
<td>Reverb</td>
</tr>
<tr>
<td>7. Mosquito buzzing</td>
<td>High</td>
<td>Medium</td>
<td>≈ 0.4 seconds, split up</td>
<td>Tight, main content around 7-8 kHz</td>
<td>Low-cut filter, slight reverb/chorus</td>
</tr>
<tr>
<td>8. Pterodactyl scream</td>
<td>Low</td>
<td>Low</td>
<td>≈ 3 seconds</td>
<td>Medium, peaks at 3 kHz and &gt; 10 kHz</td>
<td>Echoes (after each screech)</td>
</tr>
<tr>
<td>9. Monster snarling</td>
<td>Low</td>
<td>Medium</td>
<td>≈ 3 seconds</td>
<td>Medium, peaks moving from 50-500 Hz</td>
<td>Slight reverb</td>
</tr>
<tr>
<td>10. Butterfly flopping</td>
<td>Low</td>
<td>Low</td>
<td>Medium, peaks at 80 and 90 Hz</td>
<td>Slight reverb</td>
<td></td>
</tr>
</tbody>
</table>

Table 1: The first ten sound file levels in the game.
My intention with focus group session 2 was to test the how efficient the [earplug-~] filter was for the participants, to let the participants test the navigation to see if it was intuitive, and to get an understanding of the learning curve for blind users when using the iPhone’s accelerometer to manipulate sounds in a 3D space. I also wanted the participants to test some of the created game levels in order to find out whether or not they were too difficult, or of low sound quality.

### 5.3.1 Qualitative data analysis

Similar to focus group session 1, the qualitative data for this focus group consisted of the responses from the participants, which were gathered and analysed similarly as in focus group session 1. The qualitative data was also coupled with notes on how the participants achieved in various tasks.

### 5.3.2 Results

First, the participants were tested for whether or not they perceived the HRTF-filtering as expected. For simultaneous listening, two pairs of in-ear earphones were used (Pioneer-CL711-G, frequency range 8Hz-22kHz). Then, the sound of a looping drum-set was positioned and moved forward, backwards and around the listener using a graphical interface, in order to test for front/back-confusion. The participants were told to point in the direction of the sound source when being asked of its location. This was done ten times unless an error was made. Both participants pointed in the correct direction all ten times, and expressed that they had a clear perception of the position at all times. Following this, the iPhone navigation was explained and then practiced with the same drum loop for about one minute each. In order to get an indication of the efficiency of the capturing time, the sound was repositioned at the maximum distance and random azimuth location, and the participants were asked to capture the sound as quickly as possible. As explained in [navigate_turn_and_walk_around], the sound could only be captured within a range of 2% of the total radius (width), or the maximum distance to the sound object.
(i.e. the sound had to be almost exactly straight in front or behind the listener in order to be captured in a straight run). During the navigation, hand and head movements were observed, and task completion time was registered. P2’s hand movements were calm and sequential, while P1’s were more rolling, as if using a joystick. P2 captured the sound within 73 seconds, while P1 needed more than 120 seconds. Both of the participants held their heads still. When letting the participants test level 15, they both managed to capture the sound within one minute, and they stated that the noise imposed by the filter did not disturb them at all. However, when testing levels 18 and 19, none of them managed to capture the sound. Both P1 and P2 could hear the frequency in level 15 and 18, but only P1 could hear the frequency in level 16 and 19. The main reason why none of them could capture the sounds in level 18 and 19 seemed to be that these levels caused too much confusion, and it was simply impossible for them to know where the sound was. We therefore concluded that level 18 and 19 had to be removed, and that level 16’s upper frequency limit had to be increased so that P2 could perceive it. Furthermore, the participants felt that levels 4 and 8 did not sound real enough, and that the sounds in these levels had to be changed. Also, the way the navigation functioned at the time, cheating was possible. Therefore, after some further open discussion, we came to the following additional conclusions:

1. *The iPhone should be held horizontally* using two hands, as opposed to using one hand vertically, both as it promotes easier hand movements, and because the VoiceOver function alerts the user already in the game menu that this is the way the iPhone should be held, making it a more intuitive game control.

2. *There should be a practice option in the menu*, allowing the user to train their navigation during an unlimited period of time with a complex and easily heard sound.

3. *The iPhone should react faster when leaning*, and *the sound’s movement should accelerate when leaning the iPhone more*, to promote easier captures at close distances. Also, *invisible walls have to be created*, such that cheating becomes impossible.

4. *The game levels should vary in time*, with longer times at the first levels and shorter times as the game progresses to promote a sense of confidence in the player.

5. *The sound should decrease more in intensity at further distances* to give a greater illusion of externalization.

6. Apart from the notification sound alerting the player that the sound has been captured, *further auditory icons should be included* (e.g. countdown to the next level, game over and game completed alert, new highscore alert, new time record alert, as well as dynamic auditory icons representing the score amounts for each level).

### 5.3.3 Post-development

#### 5.3.3.1 Creating the second Lo-Fi menu prototype

The second Lo-Fi menu prototype stuck horizontally, thus forcing the player to hold it with two hands. Additionally, the practice option was included (see Figure 20).

![Figure 20: The second Lo-Fi prototype of the menu, stuck horizontally and including the practice option.](image-url)
5.3.3.2 Optimizing the Hi-Fi navigation prototype

Apart from switching the input accelerometer data for horizontal navigation (i.e. x was switched to y and vice versa), the navigation was also optimized to remove stack overflow errors, to react faster, and accelerate when leaning the iPhone more, all of which was done by re-programming the [turn] abstraction object (see Figure 21).

![Diagram of the [turn] object.](image)

From the incoming y values of the smartphone, 180 degrees is the zero-position (i.e. when the smartphone is held flat horizontally, nothing happens, where leaning it 90 degrees to the right corresponds to 0 degrees, and leaning it 90 degrees to the left corresponds to 360 degrees). When the smartphone leans to either side, the counter is activated, sending the degrees as output for the [navigate_turn_and_walk_around] object.

**Upgrade 1 (sensitivity and acceleration):** During focus group 2, it was clear that the sound had to react faster when leaning the smartphone. We also came to the conclusion that it might be beneficial to be able to accelerate the sound depending on how much the player leans the smartphone. Therefore, I upgraded the [turn] object, such that it instead reacts to six different positions (three on each side of the zero-position). The counter is triggered earlier for each side (240 degrees as opposed to 300 degrees for spinning clockwise when leaning to the left, and 120 degrees as opposed to 60 degrees for spinning counter-clockwise when leaning to the right). Each side also has three different acceleration modes, where leaning the iPhone 300 degrees (left) and 60 degrees (right) makes the sound spin twice as fast as the initial speed, and leaning it 340 degrees (left) and 20 degrees (right) makes the sound spin three times as fast as the initial speed.

**Upgrade 2 (error correction):** The second outlet was also changed so that it outputs a bang (activator) when the smartphone is held at the left/right-angels just before those activating the sound object to spin (instead of outputing a bang at the same time as the sound object begins to spin). This upgrade eliminated the prior stack overflow error message in the [navigate_turn_and_walk_around] object, which was due to the hot outlet of the [poltocar] object being fed into the hot inlet of the [cartopol] object (and vice versa) at the same time, causing a stack overflow.

5.3.3.3 Eliminating the possibility to cheat

Furthermore, by the way the navigation functioned at the time of focus group 2, sounds that moved beyond the maximum distance at the front or back appeared at the opposite side (such as in the Snake 2 game for old telephones). This was bad, not only because it promoted front/back-
confusion (e.g. a sound at the maximum distance in front and slightly to the right of the player suddenly appearing at the maximum distance behind and slightly to the right of the player might be perceived as still being in front and to the right), but also because it promoted cheating. The reason for this was that when the player leaned the iPhone either to the right and left while at the same time leaning it front or back, this led to the sound spinning around the player while getting closer and closer, which could lead to a perfect highscore with minimal effort (during the focus group, the participants were informed about this so they did not cheat while navigating). To eliminate these problems, *invisible walls* were created, which was done by reprogramming the [walk] abstraction object. As the built-in [counter] object in Pd did not support setting a maximum or minimum value at which incoming values would not affect the counter, I built my own [counter] abstraction object that did this, thus stopping sounds when they reached the maximum distance in front or behind the listener when walking.

5.3.3.4 Optimizing distance

As opposed to having a linear sound level decrease depending on distance (which of course is completely unnatural), I re-programmed the [SoundDistance] abstraction object, such that for every doubling of the distance to the sound object, the perceived sound level (loudness) is halved (decreases by -6 dB, as in real life), where the formula is \( L = \log(1/r) \), see Figure 22. I also set the maximum and minimum values for the sound level to prevent unpleasant loudness experiences. Distance and direction are thus handled separately (i.e. they are not interpolated).

![Figure 22: The object [SoundDistance] was created to give a proper distance simulation.](image)

**[SoundDistance]**

This object calculates the loudness for the sound object, and makes the loudness vary depending on how close the sound object is to the player. The inlet is the radius (ranging from 0-1), which is converted to the distance to the sound object (in meters).

The maximum loudness is reached at 0.37 meters, as making the loudness increase more at closer distances would lead to very high and unpleasant loudness experiences for the player. The minimum loudness (i.e. at the maximum distance to the sound object), is \( r = 0.125 \), corresponding to 8 meters.

5.3.3.5 Optimizing the game levels

*Illusions:* As mentioned, level 18 and 19 caused too much confusion. The reason for this was that the frequency changed, such that it (in level 18) went from 5 kHz at a distance of 8 meters to 20 kHz at the capturing distance, and in level 20, it went from 50 Hz at a distance of 8 meters to 20 Hz at the capturing distance. In the beginning, I thought of this simply as a good way of making truly impossible levels. However, when testing the levels with the participants, I realized that the reason why they were so impossible was not because the tone reaches our natural hearing limitations, but rather that *illusions* are created when communicating meta-level information in this way. The illusions are created when some quality in the sound (in this case the frequency) is altered in such a way that, even when there is a significant boost in the sound level the closer the sound comes (here three times as loud, or 18 dB), it is still perceived as...
loosing its intensity the closer it comes. In other words, the sound level increase at closer distances is not compensation enough for the loss in perception caused by alterations in frequency, and the sound is perceived as becoming weaker and moving further away from the listener, when in fact, it is approaching the listener and increasing in sound level (see the section on absolute thresholds for monaural and binaural sounds).

Of course, this makes navigation completely impossible, as the listener will never be able to make distance judgements. Therefore, even if these levels had to be taken away, the discovery of the illusions they created was important for understanding how meta-level communication should be used, and also how it should not be used in 3D-audio games.

**Alterations and new levels:** As mentioned, levels 18 and 19 were removed. However, as the illusions were not present on levels 15 and 16, these levels were kept and regarded as the most difficult (yet still possible to complete) levels in the game, but due to P2’s inability to hear the upper frequency limit of 50 Hz on level 16, it was increased to 150 Hz. Level 15 and 16 therefore became the new levels 19 and 20. Furthermore, levels 4 (“Fly buzzing”) and 8 (“Pterodactyl scream”) had been regarded as of low quality, and were therefore replaced by the new and better-sounding levels “Bee in jar” (level 4), and “Pterodactyl screech” (level 8).

Due to the fact that I only had 19 levels to begin with, and that two levels had to be removed, I therefore added three additional levels. As the sound files had been highly appreciated by the participants, and the synthesized levels seemed to cover a good difficulty range, I decided to add three more sound file levels, thus making it a total of 13 sound file levels and 7 synthesized levels. All the changes and added levels can be viewed in Table 3.

<table>
<thead>
<tr>
<th>New levels</th>
<th>Realness</th>
<th>Oscillation</th>
<th>Silence</th>
<th>Frequency range</th>
<th>Effects</th>
</tr>
</thead>
<tbody>
<tr>
<td>4. Bee in jar (replacing Fly buzzing)</td>
<td>Medium</td>
<td>High</td>
<td>≈ 0.2 seconds</td>
<td>Wide, main content around 100-400 Hz</td>
<td>Slight reverb, chorus</td>
</tr>
<tr>
<td>8. Pterodactyl screech (replacing Pterodactyl scream)</td>
<td>Low</td>
<td>Low</td>
<td>≈ 3 seconds</td>
<td>Medium, peaks at 3 kHz and &gt; 10 kHz</td>
<td>Echo (after the initial screech)</td>
</tr>
<tr>
<td>15. Monster walking (replacing sine sweep 20-5 kHz)</td>
<td>Low</td>
<td>Low</td>
<td>≈ 5 seconds</td>
<td>Medium, main content &lt; 100 Hz</td>
<td>Slight reverb</td>
</tr>
<tr>
<td>16. UFO (replacing sine sweep 20-50 Hz)</td>
<td>Low</td>
<td>Medium</td>
<td>≈ 1.5 seconds</td>
<td>Medium, two peaks at 700-800 Hz and 2 kHz</td>
<td>Dynamic EQ (band-pass filter), slight chorus, reverb</td>
</tr>
<tr>
<td>18. Shotgun (replacing sine sweep 5-20 kHz)</td>
<td>Medium</td>
<td>High</td>
<td>≈ 5 seconds</td>
<td>Very wide</td>
<td>Echo with long decay (3.97 seconds)</td>
</tr>
</tbody>
</table>

Table 3: The changes to existing levels and levels having been added after focus group 2.
**Method**

**Level times:** In the focus group we came to the conclusion that instead of having the same time constraints for each level (initially 20 seconds), it would be better to have longer times for the first levels in the game, as this would give the player some time to adjust to the controls, and also promote confidence in the player. Therefore, I decided to use 20 seconds for the final level, and increase the time limit for each level prior to that level with the following linear function:

\[
Y_{\text{(new level time)}} = 20 + (20 - X_{\text{(level number)}}) \times 3
\]

This equals adding three seconds to every level prior to the final level until the first level is reached (i.e. the first level has a time limit of 20 + 19*3 = 77 seconds).

**5.3.3.6 Adding auditory icons and sonification**

The final stage of the post-development for focus group 2 was adding auditory icons in order to convey confirmatory sounds related to player progress (e.g. countdown to the next level, new highscore alert, new time record alert, game over and game completed alert), and using sonification to map abstract data into non-speech sounds in order to represent dynamic data changes (e.g. representing score amounts for each level), which was done by creating the new abstraction object `[activate_auditory_icons]`, see Figure 23.

![Figure 23: The abstraction object [activate_auditory_icons], activating sonification and auditory icons in the game.](image)

[activate_auditory_icons]

This object activates the various static and dynamic auditory icons in the game.

The "level complete" sound is a confirmatory auditory icon (top left corner), notifying the player that the level has been completed (activated when the player is within the score radius).

The level highscore sounds are dynamic (i.e. they give the player a sense of how many points they achieved, where a shorter sound indicates a low score and a longer sound indicates a higher score), and depend on how fast the level was completed in relation to that particular level’s completion time. They come in four different categories, lowest, low, high, highest highscore (top right corner).
The "level highscore" sounds in turn activate the "countdown" sound, prior to the next level beginning, sending the [activate_auditory_icons] object’s only output, which repositions the sound and adjusts the loudness level for the next game level.

The "game over" sound and the "game completed" sound are activated when the game is over (i.e. a level was not completed within that particular level’s time frame, or if the whole game has been completed).

The "game over" sound and the "game completed" sound in turn activate the "new highscore" sound, but only if there is a new highscore record (i.e. the sound is heard either when the player dies, or when they complete the game).

The "game completed" sound and the "new highscore" sound then activate the "new time record" sound, with the condition that the whole game has been completed, and that there is a new highscore record.

5.4 Focus group session 3

During this focus group session, the participants tested the re-programmed navigation to see if it was more efficient and intuitive, the current version of the game featuring 20 levels varying in difficulty and time, as well the static and dynamic auditory icons.

5.4.1 Qualitative data analysis

Similar to focus group sessions 1 and 2, the qualitative data for this focus group session consisted of the responses from the participants, which were gathered and analysed similarly as in focus group sessions 1 and 2. The qualitative data was also coupled with notes on how efficiently the participants achieved in various tasks, similar to focus group session 2.

5.4.2 Results

The participants first tested the new navigation as in focus group session 2, and they felt that the controls were more intuitive this time – “You have to get used to holding it with two hands, but it is a more logical way to be working with it” (P1). However, even though they felt that the controls were more intuitive, they still had difficulties in capturing the sound – “I felt that when I had centred the sound so I knew that I had to go forward or backwards to capture it, the sound still moved past me” (P2). As their responses indicated that the score radius might be too narrow, it was broadened from 2% to 5%, which led to drastic improvements, and both of them captured the sound immediately. When testing the game for the first time, they both came to the fifth level. They thought that the auditory icons were intuitive – “Yeah, this will definitely work, you get immediate feedback” (P1), and they seemed to be highly immersed in the game – “I only thought about the game. I really felt that I was only the game, it was the one and only thing that mattered” (P2).

However, as the placement of the levels was based on my initial subjective impression of their independent difficulty levels, as well as the fact that both had expressed that the level difficulties varied a lot in the first five sound file levels, they were allowed to play some of the later synthesized levels in the game (note that these levels had shorter times). When beginning at level 11, they both managed to reach level 13. Interestingly, they both accomplished the same amount of levels in both tests, and the atmosphere was clearly competitive. Both P1 and P2 also thought that the synthesized levels were just as fun as the sound file levels – “I thought they were great. I didn’t think at all about whether the sounds were more realistic or not. I was just focused on capturing them. The only thing I paid attention to was if they were easy to hear or not” (P2). Because of this, and that the synthesized levels seemed to be much easier than even the first sound file levels in the game, we came to the conclusion that the levels should be mixed. When letting them test levels 19 and 20 (levels 15 and 16 in the second focus group session), they thought that the concept of beginning the level with silence (i.e. the sound had to be hunted in order to be perceived) was fine, as the auditory icons communicated that the level had begun – “As long as you have the countdown you know that something has started, and
when you can’t hear the sound you know that you have to hunt it down” (P1). The main conclusions from this focus group were therefore:

1. **Further usability tests were necessary in order to determine the correct level order.**
2. **The score radius should remain at 5%, or be globally adjustable.**
3. Both participants also mentioned that **the time decrease in every level should be mentioned either in the instructions**, as this knowledge would affect the excitement and immersion in the game.

### 5.4.3 Post-development

#### 5.4.3.1 Creating the Lo-Fi prototype of the game instructions

The final Lo-Fi prototype of what the game might look like on the iPhone covered the instructions, shown as text on the iPhone (as concluded in focus group 1), and mentioning the time restrictions for each level (as concluded in focus group 3). Furthermore information related to the game, as well as how to hold the iPhone was also included, see **Figure 24**.

![Game Instructions](image)

**Figure 24**: The game menu instructions shown as text, mentioning the time restriction.

#### 5.4.3.2 Making the score radius globally adjustable

The final post-development of the game was making the score radius globally adjustable in the main [SOUND_HUNER] game object, including four difficulty levels; “easy” (score radius 7%), “normal” (score radius 5%), “difficult” (score radius 3%) and “impossible!” (score radius 2%). For the usability tests, the score radius was kept at 5%, as this seemed to be a good mean difficulty level.
5.5 Part 2 – Testing and evaluating Sound Hunter

5.6 Usability tests

The main reason for conducting the usability tests was to determine the correct game level order, but also to see how well the participants perceived the HRTF filtering and how efficiently they could use it interactively. Furthermore, I wanted to find out which levels were the most difficult and why they were considered difficult, to see if the participants improved in the game over time, to make observations of the participants movements and interactions, as well as to get qualitative feedback connected to user satisfaction and the game in general.

5.6.1 Participants

Eight participants took part in the usability tests:

Participant 1 (P1): Female, 24 years old, sighted with no documented hearing impairments. She plays computer games no and then, but never audio-only games. She knew about and had passive experience with binaural sounds or HRTF filtering (e.g. the YouTube-video Virtual barber shop57), and uses an iPhone.

Participant 2 (P2): Male, 52 years old, legally blind since ten years, and had a documented hearing impairment on the right ear. He never plays computer games, but had heard of “dummy head stereo” and had little passive experience with it (e.g. Virtual barber shop), and uses an iPhone.

Participant 3 (P3): Female, 50 years old, total blindness since birth, and no hearing impairments. She had played “Shades of Doom”, had heard of “dummy head stereo”, had little passive experience with it (e.g. Virtual barber shop), and uses an iPhone.

Participant 4 (P4): Male, 16 years old, sighted with no hearing impairments. He plays a lot of computer games and has had the occasional experience with an audio-only game. He did not know of the term binaural sounds or HRTF, but had experienced 3D sound, both passively and interactively through 7.1 headphones (e.g. virtual barber shop and playing Unreal Tournament 358), and uses a Windows Phone.

Participant 5 (P5): Male, 15 years old, sighted, and no hearing impairments. He plays a lot of computer games, but had no experience with audio-only games. He did not know of the term binaural sounds or HRTF, but had experienced 3D sound, both passively and interactively through 7.1 headphones (e.g. virtual barber shop and playing Unreal Tournament 3), and uses an iPhone.

Participant 6 (P6): Male, 23 years old, born with total blindness with no hearing impairments. He plays a lot of audio-only games, and also develops them through his company Blastbay Studios59. He is knowledgeable and experienced with the term binaural sounds or HRTF, both passively and interactively (virtual barbershop, and through his own binaural microphones). Furthermore, he has played all of the commercially available 3D-audio games mentioned in this thesis, such as “Shades of Doom and Terraformers. He only owns and uses his home telephone.

Participant 7 (P7): Male, 24 years old, sighted with no hearing impairments. He plays computer games every now and then, but had no experience with audio-only games. He is knowledgeable and experienced with the term binaural sounds or HRTF, but only passively through headphones (e.g. virtual barbershop and a course at KTH), and uses an iPhone.

57 Virtual barber shop, YouTube video, http://www.youtube.com/watch?v=1UDTlvagjiA
58 Unreal Tournament 3 [support for 7.1 headphones], http://www.unrealtournament.com/uk/index.html
59 Blastbay Studios, main website, http://www.blastbay.com/
Participant 8 (P8): Female, 19 years old, legally blind since early childhood (has memories of seeing but not of what she saw), no documented hearing impairments, and was currently being investigated for ADHD. She has played some of P6’s audio-only games, but has no other familiarities or experiences with the terms binaural sounds or HRTF, and uses an iPhone. The participant selection was conducted in Stockholm, Sweden, through the help of the company Funka Nu, as well as reaching out to various organizations and associations for older and younger blind people, such as Synskadades Riksförbund and Unga Synskadade Stockholm.

5.6.2 Procedure
After explaining how the game worked, the participants’ abilities to perceive the HRTF filtering was tested similarly as in focus group 2, where the sound of a drum-set was positioned and moved around in a 3D space, and the participants were asked to point in the direction of the sound source. This was done five times unless an error was made, in which case they were tested further, mainly to conclude whether or not the participant had an efficient perception of sounds being in front and behind them, and to get a baseline measurement of HRTF adaptation. The participants were then familiarized with the navigation, and were allowed to practice as long as they wanted. When they captured the sound, it was replaced to let them capture it again until they felt comfortable enough to start playing the game. Each participant played through the entire game including all 20 levels five times each. The time restriction had been removed, thus allowing the participants to play each level as long as they needed in order to capture the sound. Automatic data-collection was used to retrieve the level times in order to quantitatively analyse efficiency. While simultaneously listening to the participants playing the game, hand and head movements were also observed (to see if the participants compensated by moving their heads and how efficiently they could use their hand movements to navigate), as well as (if necessary) how many tips had to be given on each level in order for it to be completed. After the game had been played through, a semi-structured interview was conducted with each participant in order to obtain additional information related to player satisfaction, the game in general, and possible improvements.

For reasons being explained in the qualitative data analysis, P8 unfortunately had to be excluded from the quantitative data analysis, but she was still part of the qualitative data analysis. The quantitative data analysis was therefore performed for participants P1-P7.

5.6.3 Quantitative data analysis
5.6.3.1 Determining level order
The difficulty of each game level was determined by analysing the completion times, as well as the amount of tips that had to be given for each level. The level completion times were retrieved and saved separately using automatic data-collection during each gaming session.

5.6.3.2 Tips
The tips were categorized into amount, participant, game level, as well as which session they belonged to, see Table 4. A tip was given if the participant communicated uncertainty connected to something previously being explained (e.g. how the navigation worked), or if 90 seconds had passed without level completion. The given tip was always connected to the navigation (e.g. “try leaning forward”), regardless of what caused the uncertainty, such as a certain level being more difficult, front/back-confusion, or if the participant could not hear the sound object at a certain distance (being the case for levels 19 and 20, where the sound only becomes audible at closer distances).
<table>
<thead>
<tr>
<th>Level (original order)</th>
<th>Amount of tips required for level completion</th>
<th>Session</th>
<th>P1</th>
<th>P2</th>
<th>P3</th>
<th>P4</th>
<th>P5</th>
<th>P6</th>
<th>P7</th>
<th>Level total</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Drum-set</td>
<td></td>
<td></td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>2. Spooky chains</td>
<td></td>
<td>S1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>3. Baby talking</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4. Bee in jar</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5. Spooky water drops</td>
<td></td>
<td>S1</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>5</td>
</tr>
<tr>
<td>6. Raven</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7. Mosquito buzzing</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8. Pterodactyl screech</td>
<td></td>
<td>S1</td>
<td>3</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>9. Monster snarling</td>
<td></td>
<td>S1</td>
<td>1</td>
<td>3</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>10. Butterfly flopping</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11. Square click CoF 1-9Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12. Square click CoF 9-1Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13. Square sweep CoF 2-262Hz</td>
<td></td>
<td>S1</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>14. Square sweep CoF 262-2Hz</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15. Monster walking</td>
<td></td>
<td>S1</td>
<td>4</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>16. Ufo space ship</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17. Temporal triangular noise</td>
<td></td>
<td>S1</td>
<td>1</td>
<td>5</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>6</td>
</tr>
<tr>
<td>18. Shotgun blast</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19. Sine sweep 20kHz – 5kHz</td>
<td></td>
<td>S1-5</td>
<td>1</td>
<td>7</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>1</td>
<td>1</td>
<td>14</td>
</tr>
<tr>
<td>20. Sine sweep 20Hz-150Hz</td>
<td></td>
<td>S1-5</td>
<td>4</td>
<td>9</td>
<td></td>
<td></td>
<td>1</td>
<td>1</td>
<td></td>
<td>15</td>
</tr>
</tbody>
</table>

Table 4: Number of tips given to participants on different levels and sessions.

Almost all of the tips were given in the first gaming session, with exception for P2 who had a slight hearing impairment, and therefore was only able to hear game level 19 at a very close distances, meaning that many tips were required in order for him to complete it (where the average completion time was over 100 seconds). Also, as he could not hear level 20 at all, I decided to put 100 seconds on both level 19 and level 20 for P2 for the quantitative data analysis. The rest of the participants’ completion times were kept as they were.

In general, the tips seemed to be connected mainly to the participants’ individual learning curves related to the navigation. They did, however, also seem to relate to the difficulties of each level as expected (i.e. the order was not logical), which was supported when analysing the level completion times, as well as by the qualitative data analysis.

5.6.3.3 Level completion times

The new level order (i.e. level completion efficiency) was determined by comparing the level completion times using a One-way Repeated Measures Anova in SPSS. The within-level difference in level completion time was significant F(5.322, 180.958)=14.31, p<0.01, and 29.6 per cent of the total variance could be explained by the completion time (using the Greenhouse-Geisser correction, sphericity assumed). As expected, the levels’ difficulties did not match the order in which they had been placed. The original levels 5 and 14 were the only levels having
the same mean completion time, but as the standard error was slightly higher for level 5, as well as the fact that more tips had to be given for this level, it was considered more difficult than level 14. The results from the analysis, and the new level order, are shown below in Table 5. An overview of their comparable difficulties is given in Figure 25.

<table>
<thead>
<tr>
<th>Original Level</th>
<th>New Level</th>
<th>Tips</th>
<th>Mean level completion time (s)</th>
<th>Std. Error</th>
<th>95% Confidence Interval</th>
</tr>
</thead>
<tbody>
<tr>
<td>L.11</td>
<td>L.1</td>
<td>8,257</td>
<td>0.874</td>
<td>6,481</td>
<td>10,033</td>
</tr>
<tr>
<td>L.12</td>
<td>L.2</td>
<td>9,886</td>
<td>1.043</td>
<td>7,767</td>
<td>12,005</td>
</tr>
<tr>
<td>L.13</td>
<td>L.3</td>
<td>1</td>
<td>11</td>
<td>8,457</td>
<td>13,543</td>
</tr>
<tr>
<td>L.10</td>
<td>L.4</td>
<td>11,886</td>
<td>1.225</td>
<td>9,396</td>
<td>14,375</td>
</tr>
<tr>
<td>L.10</td>
<td>L.5</td>
<td>12,714</td>
<td>1.889</td>
<td>8,876</td>
<td>16,553</td>
</tr>
<tr>
<td>L.3</td>
<td>L.6</td>
<td>12,829</td>
<td>1.043</td>
<td>10,709</td>
<td>14,949</td>
</tr>
<tr>
<td>L.4</td>
<td>L.7</td>
<td>13,857</td>
<td>1.399</td>
<td>11,013</td>
<td>16,701</td>
</tr>
<tr>
<td>L.14</td>
<td>L.8</td>
<td>15,657</td>
<td>2.944</td>
<td>9,674</td>
<td>21,641</td>
</tr>
<tr>
<td>L.5</td>
<td>L.9</td>
<td>15,657</td>
<td>3.676</td>
<td>8,186</td>
<td>23,128</td>
</tr>
<tr>
<td>L.2</td>
<td>L.10</td>
<td>16,143</td>
<td>3.025</td>
<td>9,996</td>
<td>22,29</td>
</tr>
<tr>
<td>L.16</td>
<td>L.11</td>
<td>16,771</td>
<td>2.6</td>
<td>11,488</td>
<td>22,055</td>
</tr>
<tr>
<td>L.6</td>
<td>L.12</td>
<td>18,657</td>
<td>2.734</td>
<td>13,1</td>
<td>24,214</td>
</tr>
<tr>
<td>L.7</td>
<td>L.13</td>
<td>18,914</td>
<td>2.701</td>
<td>13,426</td>
<td>24,402</td>
</tr>
<tr>
<td>L.18</td>
<td>L.14</td>
<td>21,914</td>
<td>2.641</td>
<td>16,548</td>
<td>27,281</td>
</tr>
<tr>
<td>L.17</td>
<td>L.15</td>
<td>24,814</td>
<td>4.469</td>
<td>15,732</td>
<td>33,896</td>
</tr>
<tr>
<td>L.8</td>
<td>L.16</td>
<td>26,857</td>
<td>3.988</td>
<td>18,752</td>
<td>34,962</td>
</tr>
<tr>
<td>L.9</td>
<td>L.17</td>
<td>27,857</td>
<td>3.606</td>
<td>20,529</td>
<td>35,185</td>
</tr>
<tr>
<td>L.15</td>
<td>L.18</td>
<td>30,257</td>
<td>5.114</td>
<td>19,864</td>
<td>40,65</td>
</tr>
<tr>
<td>L.19</td>
<td>L.19</td>
<td>47,243</td>
<td>6.537</td>
<td>33,957</td>
<td>60,529</td>
</tr>
<tr>
<td>L.20</td>
<td>L.20</td>
<td>55,143</td>
<td>7.572</td>
<td>39,755</td>
<td>70,531</td>
</tr>
</tbody>
</table>

Table 5: The new level order revealed from usability testing using automatic data-collection of level times.
As seen in the figure, the easiest levels were level 11, 12 and 13 (new levels 1, 2, 3), which were some of the levels communicating additional meta-level information (in this case the cut-off frequency, altering the complexity of the sound (see levels 11-13 in the post-development section of focus group 1 on Planning Sound Hunter’s game levels). As all of the synthesized levels (except one, new level 15) communicated meta-level information, where the other two (new levels 19 and 20) were the most difficult, I will refer to the new levels 1, 2, and 3 as the easier synthesized levels in the remainder of the thesis.

Furthermore, the mean level times were all under one minute, indicating that the time restrictions calculated for each level (see the post-development section in focus group 2 on Optimizing the game levels), might be too long, and the game too easy. Optimizing the level times therefore became one of the topics for the final evaluation.

5.6.3.4 Analysis of practice effect and player improvement

The practice effect between gaming sessions (i.e. session completion efficiency) was analysed in SPSS using a One-way Repeated Measures Anova (see Figure 26), where the unit of measurement was the level times for each gaming session. The within-subjects effects were significant F(2.441, 339.332)=11.572, p < 0.01), and 10.6 per cent of the total variance in game completion time could be explained by the different gaming sessions (using the Greenhouse-Geisser correction, sphericity assumed). The differences between gaming session one and two t(139)=3.90, p<0.01, one and three t(139)=4.46, p<0.01, one and four t(139)=3.62, p<0.01, and one and five t(139)=4.41, p<0.01 were significant. All other differences and pairwise comparisons between gaming sessions were non-significant.

There are many possible explanations for the significant difference between the first and second gaming session. For example, as the participants had already played through all the levels in the first session, the surprise effect was lower, and it can therefore be argued that they had a better
idea of what they could expect in the second session (even though all of the sounds were randomly placed in each session). Another explanation could be the fact that the participants learnt how to use the iPhone to manipulate the sound movement, for which the learning curve varied a lot between participants. However, as indicated by the number of tips given, all of the participants seemed to have learnt the controls after gaming session one. Furthermore, the differences could be explained by the fact that the participants seemed to pay more attention to their hearing, which can be explained mainly by the fact that the levels differed in difficulty, where some levels needed more focus than others, but it could also be an indication that the participants were adapting to the generalized HRTF, training to perceive sounds through it as if it were their own. Finally, the sample size for the usability tests was very small, making it difficult to draw conclusions meant for a greater population.

Figure 26: Mean differences in seconds between the completion times for each played session.

Even though no significant differences were found between the gaming sessions two to five during the usability tests (most likely due to the low sample size), the results still show trends in how the players might have improved in the game over time. The number of tips indicates that all of the participants had a better understanding of the game and its controls after the first session. It is therefore possible that the time decrease between sessions two and three could be explained by the fact that the participants seemed to be focusing more on their hearing. This is also reflected in the difference between sessions three and four, as most of the participants stated that they were becoming tired during the fourth or fifth session (see observations section of the qualitative data analysis). As mentioned above, this could also be because of the adaptation to a foreign HRTF under virtual interactive circumstances. A possible explanation for the improvement between sessions four and five could be that the participants felt a sense of relief, as they knew that they soon would be finished with the test. After all, playing through the
entire game five times without any time constraints or imposed challenges except locating a sound object may become quite tiresome after four to five sessions (equalling roughly one hour of uninterrupted gaming). Most of the results and trends from the quantitative data analysis were confirmed in the qualitative data analysis, as well as in the final qualitative expert evaluation.

5.6.4 Qualitative data analysis

The qualitative data analysis of the usability tests was divided into two parts, the observations I made, as well as the qualitative feedback I got from the participants during the post-game interview. As mentioned, P8 was included in this analysis, but excluded from the quantitative data analysis, the reason for which will be explained in this section.

5.6.4.1 Observations

To get an understanding of how the participants behaved when playing the game, special attention was paid to their hand and head movements (to see if they tried to compensate by leaning their heads and how they utilized the navigation through hand movements), their passive and interactive front/back-confusion, as well as their overall improvement and immersion in the game (see Figure 27 and Figure 28).

**HRTF perception:** Before letting the participants navigate, I performed a test similar to that in focus group 2, where I positioned a sound in a 3D-space (drum-set), and asked the participants to point in the direction of the sound source. The reason for doing this was mainly to see if the participants experienced front/back-confusion, and the test was given five times unless a wrong answer was given, in which case they were tested five more times.

All the participants except P5 and P8 pointed in the right direction all five times. P5 got 2/5 correct answers, where further tests showed that he perceived both front and back. P8 got 3/5 correct answers, where further tests showed that she perceived sounds only from the front and not behind her.

**Hand and head movements:** These movements were observed in order to get an understanding of how the participants used the controls, how long it took for them to understand how to use them, as well as how they held their heads (i.e. if they tried to compensate by leaning their heads). For the sighted participants, I also made notes on whether or not they closed their eyes. Four of the participants (P4, P5, P6, and P7) immediately understood how to use the controls, either performing slower sequential hand movements (P4, P5), or faster sequential hand movements (P6, P7). All of them held their head still throughout the session, except P5 and P6, who leaned the heads slightly forward at times. P5 never closed his eyes, P4 closed his eyes occasionally, and P7 kept his eyes closed throughout the session.

P1 and P3 span the iPhone in all directions at first, but started using more sequential movements after a while, where P1 learnt the movements quite quickly (after about two levels), held her head still at all times, and occasionally closed her eyes. P3 leaned the iPhone more backwards in the beginning, and her stance indicated that she was more used to holding the iPhone backwards and more upright. P3 turned her head slightly in the beginning, but then held it still. However, as soon as she played the easier synthetic levels, she began moving the iPhone both more sequentially, and started leaning it forward and backwards, holding her hands higher up. After the first session, she therefore seemed to have learnt the hand movements.

P2 used sequential but extremely careful hand movements in the beginning (i.e. he leaned the iPhone only very little), but after a short time (about two levels) he began leaning the iPhone more. He also held his head still at all times.

P8 held the iPhone leaning slightly backwards (as P3), and moved it heavily to the sides, rolled it around, but seldom moved it forward or backwards (even after several tips). She lost focus very quickly, and tried to hold the iPhone more as usual, turning it slightly upright with the left hand. She held her head straight, but quickly sank down, leaning her head slightly forward and lowering her shoulders. However, when playing the easier synthetic levels, there was quite a remarkable observational change. She immediately became much more engaged in the game,
sat up straight and held her arms higher (a lot like P3). During and after these levels, she also began to perform more sequential hand movements.

**Figure 27:** One of the participants playing Sound Hunter in their natural home environment.

**Front/back-confusion:** I wanted to observe whether interactive front/back-confusion differed from passive front/back confusion, and how it was linked to the navigation (i.e. if front/back-confusion caused navigation difficulties, if navigating could eliminate front/back-confusion, or if there was a problem with the navigation in general)

All participants used the navigation to eliminate front/back confusion (i.e. when experiencing front/back confusion, they leaned the iPhone to spin the sound to hear where it was, followed by capturing it), where some of the participants preferred capturing sounds from behind (P2, P3), and some of the participants preferred capturing them from the front (P1, P4, P5, P6, P7). For the participants who got lower results in the passive HRTF-perception test, the results varied. P5’s front/back-confusion during the initial test seemed to be completely eliminated when he started playing the game. It went very well from the start, and he captured the sounds both from the front and back. For P8, however, the observations were completely different. In the beginning, it seemed as if she could not hear any difference between front and back (as in the initial test), and it was extremely difficult for her to capture even a single sound (a capture averaging several minutes, or about an hour per session), but when she did capture a sound, it was always captured from the front. However, when playing the easier synthesized levels, she seemed to perceive the sounds much better in general, as well as the 3D positioning, where she captured sounds both from the front and back. She also navigated a lot better on these levels, and never gave up until she had captured the sound (averaging about half a minute in completion time).

In general, the easier synthesized levels seemed to eliminate front/back-confusion, as well as teach the players having difficulties with the controls how to use them, where the differences were clearly observable.

**Practice effect:** All of the participants except P8 navigated without any problems after the first gaming session, and four of the participants (P1, P2, P6, and P7) stated that they were becoming tired during the fourth or fifth gaming session, supporting the quantitative data analysis on practice effect. P8 had a much longer learning curve for the navigation than the rest of the
participants, and the reason for this seemed to be her inability to concentrate. Often, she gave up before capturing the sound, expressing that she wanted to do something else, such as to fetch a glass of water [I convinced her to try a bit more]. As mentioned, however, the real observational difference occurred when she played the easier synthesized levels. The reason for this was that the continuously alternating meta-level information seemed to capture and maintain her interest. Even the easier continuous levels (e.g. the drum-set) could not maintain her interest in the same way as the easier synthesized levels, and it seemed as if she learned a lot when playing these levels.

Figure 28: Close-up picture of the participant’s navigational hand movements while playing Sound Hunter.

**Immersion:** Five participants (P2, P3, P6, P7, and P8) communicated signs of immersion while playing the game. This was either done through statements (P2 – “The sounds are very fun!”), (P6 – “Now I took you, you son of a bitch!”), through body language (P3 smiled almost all the time, occasionally laughing out loud, or P8 when playing the easier synthesized levels, where it, from an observer’s point of view, looked as if she was driving a virtual vehicle), or a combination of both (P7, who showed both enjoyment through laughs in delight, but also feelings of disgust and visible shivers on the insect levels – “Uuough! I hate bees! I lose my focus on this level [Fly buzzing], it gives me shivers in my whole body”).

**Audibility:** Generally for all participants, it seemed to be slightly easier to make direction judgements at closer distances. Also, even though I had expanded the frequency range for level 20 after focus group 3, three participants (P2, P3, P8) had difficulties hearing the final levels of the game. P2, who had a documented hearing impairment on the right ear, perceived the tone in level 19, but only at a very close distance (r < 0.1). He did, however, manage to capture the sound in level 19, but only after a long time (≈ 140 seconds). Level 20 he unfortunately could not hear at all. As mentioned in the quantitative data analysis, I decided to put 100 seconds on both of these levels for P2 in the quantitative data analysis. P3 could hear level 19, but level 20 she could only hear at r ≈ 1/3. However, as she completed all the levels, no adjustments were made in the quantitative analysis. P8 stated that she had no hearing impairment, but during the test I had to increase the volume by 3 steps (=18 dB, which is over the border of becoming uncomfortable for most listeners, yet still within the range of normal hearing (see the section on Absolute thresholds for monaural and binaural listening)). With the volume increased at her
preference, I performed a simple hearing test by asking her when she started hearing the low frequency tone in level 20. She started hearing it at r=0.170 (≈1/6 of the total distance, equalling a 128 Hz tone at a distance of 1.36 meters straight in front of her). Even though I could not definitely determine whether or not she had a hearing impairment, she seemed to have much greater difficulties in hearing the sounds than the rest of the participants, despite her young age. Finally, in general, the inter-aural level differences (ILD’s, as used in level 19) seemed to be more effective cues for localization than inter-aural time differences (ITD’s) or inter-aural phase differences (IPD’s, used in level 20) for pure sine waves during virtual interactive gaming experiences. However, this cannot be said for sure, as the overall sound level was more difficult to perceive in level 20 for P2, P3, and P8.

**Other notes:** P8 never completed the usability test, simply because it took too long (two hours for two sessions, as opposed to about 40 minutes for five sessions for the rest of the participants). The results were also too biased by tips and uncompleted levels, and I therefore decided to exclude P8 from the quantitative data analysis. During the post-game interview, P8 stated that she was currently being examined for ADHD, and that she therefore always felt stressed, and had great difficulties concentrating or sitting still, which was also clearly noticeable in the observations. P8’s ability to efficiently navigate was only present on levels 11-13 (including complex distance-dependent meta-level communication).

### 5.6.4.2 Post-game interviews

The post-game interview consisted of a semi-structured questionnaire with 51 questions, covering background information, localizing sounds in the game, navigating, the character of the sounds, externalization, performance, immersion, feelings and emotions, and about the game in general. Some of the questions were of nominal character (e.g. gender, age), some were statements measured on a 7-point Likert scale (e.g. “I was confused during the game”, ranging from never, very seldom, seldom, neither seldom nor often, often, very often, always), and some were open questions (e.g. “what was the worst part of the game?”). For a complete overview of the questionnaire, I refer to the usability questionnaire in the appendix section (Swedish).

The main reasons for conducting the interviews were to find out whether or not Sound Hunter was perceived as being effective in its goals (see the method section on Effectiveness, efficiency, and satisfaction), and if the participants were satisfied with the game (keep in mind that the usability tests did not include the entire game with time constraints and auditory icons). In general, the qualitative feedback from the interviews confirmed most of the observations, and I will therefore summarize the topics having been covered, as well as add some of the most relevant quotes.

**Localizing sounds in the game:** All the participants stated that the sounds often were in the same place they thought they were, and that it was slightly easier to localize sounds from the sides than from the front and back, except P5, who thought it was easier from the front and back than from the sides (i.e. completely opposite the initial perception test). P4 stated – “It felt like you really could hear where the sound was, except that it sometimes was difficult to hear if it was in front or behind”, indicating that front/back-confusion did not seem to have an impact on the enjoyment of the game, as long as it could be corrected by navigating. Three of the participants (P2, P3, and P7), occasionally experienced sounds as being above them, but none of the participants experienced the sounds as being below them. All of the participants thought that the way the sound objects moved (i.e. excluding their character) sounded very or extremely realistic, except P8, who did not think they moved realistically at all. Furthermore, all of the participants stated that they were never, or very seldom, confused while playing the game, except P8, who stated that she was very often confused while playing, but that this depended on the game level.

**Navigating:** All of the participants except P8 felt that the navigation was easily understood, where some participants understood it immediately (P4, P5, P6), and others felt that they became better and better at navigating the more they played (P1, P2, P3, P7), where P7 stated – “I don’t think there is anything wrong with the controls. As I said before, it’s a bit of a challenge but that’s the way it should be. I think it responds well”. For P6, navigating was easy,
but he was not satisfied with the sensitivity – “how you like your controls is something very individual, and I don’t like smartphones in general, but for me it was mainly the sensitivity that I didn’t like, as I am used to more instant feedback from computer buttons. A possible solution could be if you had some kind of sensitivity slider”. He also mentioned that another possible solution could be to re-program the navigation, where the sound spins less the closer it comes.

P2 stated that he became a lot better at navigating while playing, but that he felt that the navigation was the wrong way round (i.e. he wanted the sound to spin to the left when leaning left, and move forward when leaning forward, which could explain the observations, where he always captured the sounds from behind). P8 thought the game was extremely difficult – “both the controls and because I had a really hard time hearing where it was because I couldn’t concentrate, which made me stressed. But it was not because I felt performance anxiety. It’s how it always is for me. But I thought some of the levels were much more fun, but I don’t know why…[After explaining the difference between audio loops and the synthesized levels]… In that case the easiest and most fun levels where some of those easier synthetic levels, but I also liked the baby because it sounded so real”. This supports the observations, and indicates that navigational 3D-audio games might not be suitable for people with ADHD.

The character of the sounds: All of the participants thought that the overall sound quality of the game levels was very or extremely good. However, as in focus group 2, I wanted to test whether the participants noticed imposed noise from the HRTF filter present on levels including very high frequency content (new levels 13, 16, 15, and 19, where the noise was most noticeable on level 19). Three of the participants mentioned that they had noticed the noise, where one participant (P7) neither was disturbed by the noise, nor thought it was negative for the game, one participant (P4) was not disturbed by the noise, but thought it was negative for the game, and one participant (P6) thought it was both disturbing and negative for the game.

Externalization: Five participants (P1, P3, P4, P5, and P7) experienced the sounds as being located more outside the head than inside the head (very often to all the time), one participant experienced the sounds as being located more inside the head than outside the head (P2), and two participants were unsure (P6, and P8, where P8 could not decide, and P6 stated – “Although it is difficult to answer, as the sounds in a sense always sounded as if they were outside the head, but in order to make it sound truly real you need to have early reflections and proper room simulations. I am also used to 3D audio so I know what to expect, and having proper reflections is one of those things”). Also, even though the easier synthesized levels were easier to localize and seemed to prevent front/back-confusion, four participants (P3, P6, P7, and P8) thought that they sounded less externalized than levels including real sounds. This indicates that the perception of externalization is correlated with the realness of a sound. Interestingly enough, most of these same participants also stated that the easier synthesized levels were some of the most fun levels in the game (such as P8’s statement above, P7’s statement – “Effects could make it easier, such as the water drop level, where I could follow the echo. The sounds that vibrated [easier synthesized levels] were also easier, as you could basically feel the vibrations in your whole body. There was one level that sounded like a drill [new level 1] that was really fun”, and P3’s statement – “The levels varied in difficulty. The easier ones were those that changed depending on distance [easier synthesized levels], and the sounds that you are used to hearing in real life”). This supports the conclusions from focus group 3 and the quantitative data analysis on level difficulty, indicating that achieving a high amount of realness is not as important as getting the right balance between realness and a functioning 3D-audio navigation.

Performance: All of the participants felt that they improved in the game because they remembered the game levels, where three participants (P1, P2, and P6) felt that they improved quite a lot due to this, and five participants (P3, P4, P5, P7, and P8) felt that they improved only slightly or very little due to remembering the levels, supporting the surprise effect trend in the quantitative data analysis. All of the participants (even those who were blind), except P6, felt that they focused more on their hearing more or much more than usual, and that they also trained their hearing quite a lot or very much while playing the game. However, the participants’ increased focus also seemed to be linked to the tiredness having been observed for some participants after about 30 to 40 minutes of gaming – “It felt like you had been working
Method

out, but with your ears” (P7), supporting the quantitative data analysis on practice effect. An interesting quote from P2, was – “I tried to visualize and see the sounds while playing”, indicating that he tried to combine his memories of visual cues with auditory cues in order to aid localization and performance. Unfortunately, I never asked the sighted participants whether or not they felt that they performed better when closing their eyes, but the observations indicated that they closed their eyes and held their head more upright when they experienced difficulties in the game and had to concentrate more.

Immersion: Six participants (P1, P2, P3, P4, P5, and P7) felt highly immersed in the game (they paid less attention to the world outside the game, stopped thinking about time and space outside the world of the game, and felt as if they were there in the world of the game), and two participants (P6, and P8) were more in a state of deep concentration, or flow, rather than in a state of immersion while playing the game (they paid less attention to the world outside the game, stopped thinking about time and space outside the game, but did not feel as if they were there in the world of the game).

Feelings and emotions: All the participants felt happy while playing the game, where one participant (P1) felt extremely calm, six participants (P2, P3, P4, P5, P6, and P7) felt stress as a form of competitive adrenaline rush to complete the levels as quickly as possible, and P8 felt very stressed, but neither in a good nor bad way, as this was her normal state. She also felt happy only on certain levels (the easier synthesized levels and some of the audio-loop levels, such as the talking baby).

About the game in general: The qualitative feedback from the participants indicated that the game was enjoyable and satisfactory, and all of the participants would have recommended Sound Hunter to others. The concept of 3D-audio was highly appreciated, whereas the opinion of the game idea differed between participants.

Two of the blind participants (P2 and P3), felt that it was extremely fun just being able to play a game at all, and they also pointed out its potential usage for training one’s hearing – “The best part was being able to play a game with motion as a visually impaired person, and I would recommend the game to others, even sighted people, as it shows how important our hearing is” (P2), and P3 – “The best part was that it felt very interesting. I hardly ever play computer games and I thought it was very fun. I think it is good also to train one’s hearing, and I could imagine that this could be used for all kinds of other things also. It would probably be a great game for kids with special needs”.

Other participants were more focused on the quality of the sounds and 3D rendering, such as P5 – “I thought it was easy because you could hear the sounds very well. The quality of the sounds was very good, and it was also quite easy to hear where they were”, P7 – “I thought the various sounds as well as the whole experience of the 3D sound in the game was cool and fun”, and P6 – “The 3D audio and your fantastic simulation! It was really easy to localize sounds. Other games do not give provide this. I have never heard anything like what you have done here, and I have tested everything [audio game developer], and of everything I’ve heard, you have come the closest to providing real 3D audio”.

All of the participants thought that silence was difficult, where three participants (P3, P4, P5) felt that it made them focus more on what they heard, and one participant (P7) made use of the silence during the later sessions to know when to capture the sounds. However, only two participants (P2 and P7), mentioned that silence could be negative, where P2’s motivation was that he could not hear the last levels – “The worst part was that I couldn’t hear the last levels, which made me a bit frustrated, as I felt that I really wanted to complete the game”, and P7 mentioned that it could become confusing – “It was more difficult for levels with silence. The first time I played the final levels, I didn’t understand that you had to search for the sounds, which I did the second time I played the game, so I think this could be a good thing to mention somehow in the game”. How to handle silence was therefore one of the topics for the final evaluation.

Regarding further development, two participants (P1 and P4), mentioned that the game could be improved by including more levels – “If there were new levels and new sounds I don’t think I
would have become as tired” (P1), and P4 – “I thought the best thing with the game was that it was a new type of game, and it’s a bit of a challenge also. An improvement could be to have more levels”. P1 also mentioned that including the time limit and in other ways making the sounds more difficult to capture would be fun. Ways of making the game more difficult had been created, but were not part of the usability tests (e.g. global difficulty adjustment in capture range and the time limits). However, as including more levels had not been discussed in detail, this was therefore also a further topic for the final evaluation.

A possible explanation as to why sighted people seemed to find the game unusual, yet new and fun could be that the visual cues were removed, thus forcing the player to rely only on auditory cues. This mixture of enjoyment and conflict was illustrated in one of P5’s comments – “The best thing about the game was that I’ve never played anything like it, and I would recommend it to others because of this. One thing that I thought of while playing the game that perhaps could be an improvement is if you had some kind of visual aid on the screen, like an arrow or something pointing in the direction of the sound. It shouldn’t show exactly where the sound is, but only how the player is moving, but maybe that would make it too easy, I don’t know”. However, this also illustrates the need for further directional advice, which is something that could be communicated by an auditory interface just as easily as a graphical interface, and was therefore also a topic for the final evaluation.

P6 and P7 were more oriented towards further development including virtual vehicles (also supporting the observations), where P7 stated that – “The best thing with the game was the straight run on the drill sound. Actually, I could imagine that car games could be quite fun for this kind of thing”, and P6 saw opportunities in including more advanced auditory environments, and opponents with artificial intelligence – “You should make a game where the player is in an airplane and flies around and captures things while looking out for things shooting at them, so it’s good to have more artificial intelligence also”. Ways of including more artificial intelligence was another subject during the final evaluation.

Finally, the results indicated that the time it takes to adapt to a foreign HRTF differs between passive and interactive experiences, where most of the participants adapted extremely quickly during the initial perception test, but slower when using the same HRTF interactively. This is most likely also related to the participant’s individual learning curve for the navigation (e.g. P2, who had a low score in the initial perception test, but immediately adapted when playing the game). Therefore, the participants’ individual learning curves for the navigation and adaptation to a foreign HRTF seemed to be the main cause of the differences between gaming sessions one and two in the quantitative data analysis on practice effect. Optimizing the navigation (e.g. by including a sensitivity slider or by making it possible to customize the navigation such that the player can choose their individual settings), as well as eliminating front/back-confusion, were therefore additional main topics I wanted to discuss in the final evaluation.

5.7 Final Evaluation: Sound Hunter and the future

The purpose of the final evaluation was to introduce Sound Hunter’s framework and its purpose, to present the development process and its most important results, as well as letting the participants play Sound Hunter with the changes having been made throughout the development process. After the game test, we discussed various aspects of Sound Hunter (e.g. navigation, perception) and how to improve them. Finally, we discussed how Sound Hunter might be developed further, as well as what similar types of games might be created when using the same framework.

5.7.1 Participants

For the final evaluation I needed new visually impaired participants, either with experience of audio-only games, or with at least some knowledge of development, usability, or computers in
general. The reason for this was that I wanted this session to be more expert-oriented, hopefully leading to quality feedback on the game. However, finding people with these requirements turned out to be extremely difficult. During a time period of almost a month, I reached out to various companies, organizations, and social media groups, such as Funka Nu, Synskadades Riksförbund, and Unga Synsakadade Stockholm. After having to cancel two planned sessions, with a total of four middle-aged visually impaired experts due to their inability to attend, I decided to contact P6 from the usability tests. P6 was not only an audio-only game developer with high expertise in development, computers, binaural sounds, and audio-only games, but he also gave some interesting suggestions during the usability tests, and had a genuine interest in the development of Sound Hunter. He also had total blindness since birth, and fit the intended user group. Therefore, I decided that his bias due to prior experience with the game had little influence on his ability to provide quality feedback. As P6 also took part in the usability tests, I will simply refer to him as P6 here in the final evaluation as well, in order to avoid confusion.

5.7.2 Qualitative data analysis

Similar to the focus groups and usability tests, the qualitative data consisted of participant statements being recorded using an OLYMPUS WS-450S Digital Voice Recorder in an extensive semi-structured interview. The data was then transcribed and categorized according to the predefined topics in order to find hierarchies of importance connected to game improvement and future development.

5.7.3 Results

After introducing the framework, Sound Hunter’s development process, and the additions to the game since the last time he had played it (correct level order, time adjustments for the levels according to their order, additional auditory icons, availability to globally adjust the score radius), I let P6 test the game. After the test, we began discussing various aspects relating to the further development of Sound Hunter, as well as what other games might be created by following the principles of Sound Hunter’s framework.

5.7.3.1 General aspects

iOS games: As P6 had stated in the usability tests that there are many blind iPhone users, but that hardly any of them use the iPhone to play audio-games (the main reason being the lack of audio-only games for iOS users), I began by asking him whether he thought that blind iOS users would play more audio-only games on the iPhone, given that the market for them increased. P6 was confident about this – “Yes, absolutely, the amount of games and game developers for iOS increases all the time, so the interest is definitely there. I am completely convinced that you could capture the market there if you intend to develop a fully functioning iPhone game”.

Optimizing level times: In focus group 2, we came to the conclusion that it would be better to include longer times on the first levels, in order to increase the player’s confidence. However, the usability tests seemed to indicate that the game may be too easy, as the mean completion times for all levels were under one minute. As the level times in Sound Hunter decrease linearly, I proposed that perhaps the times instead should decrease logarithmically, and steeply decrease after the first levels until they reach a limit. We came to the conclusion that this depends on whether I want to keep the same game structure, in which case the time could be more complex, and decrease steeply after the first levels, followed by perhaps increasing it a little again, and then decreasing it and so on, as this would lead to more variation in the game.

5.7.3.2 Navigation

Navigational aids: Due to differences in opinions regarding the controls, we discussed how the controls might be set to the user’s preference, as well as additional navigational aids that could be implemented. During the usability tests, P6 mentioned two proposals, where the first was to include a sensitivity slider, and the second was to re-program the navigation, where the sound spins less the closer it comes. When discussing these proposals, P6 thought that including a sensitivity slider would be enough – “There are two reasons why this is better, first of all
because it is easier to implement, and second of all because re-programming the navigation actually would be illogical, as sounds don’t work that way naturally” – (i.e. when spinning the head, the sounds around the listeners head will spin equally fast, regardless of their distance). Furthermore, I had developed two additional suggestions. The first suggestion was to include an “Options”-button in the menu, where the user would be able to completely customize the navigation to their personal preference, such as holding position (horizontal/vertical), sensitivity (reaction time), spinning direction (e.g. sound spins left as opposed to right when leaning the iPhone left), and acceleration (number of acceleration steps). The second suggestion had to do with the fact that the easier synthetic levels, both during the focus groups and usability tests, seemed to eliminate front/back confusion, and also teach the player how to use the controls. Therefore, the suggestion was to always include these levels in the beginning of the game. Regarding the total customizing option, P6 thought this was unnecessary – “You often want a default setting that you know works well and that also is generalizable for a large number of people”. My observations, both from focus groups and usability tests, indicated that blind people using iPhones seemed to hold them more upright and vertically (often close to the ear while scanning the icons on the iPhone with the fingers to hear the VoiceOver function). However, as the participants seemed to adapt very well to the controls while playing, we came to the conclusion that including further customizing options were unnecessary, especially as the VoiceOver function informs the user of the position, and that the instructions inform the player how to navigate. Regarding the second suggestion, we concluded that this could be a good idea, but that it would be better to include these sounds in the practice section of the game.

**Eliminating front/back-confusion:** During the usability tests, one of the sighted participants, (P5) pointed out that a further development of the game could be to include a graphical interface where an arrow points in the direction of the sound. As finding ways to reduce front/back-confusion is always good, I proposed some suggestions of making an auditory interfaced version of P5’s proposal. The first suggestion was to display text on the iPhone, giving advice to the player, which could be activated by an auditory icon and read by the iPhone’s VoiceOver function by holding a finger on the screen (e.g. “front-left!”). The second suggestion was to include short sound bounces directly opposite the target sound in the azimuth plane and at a lower intensity, thus informing the player that the stronger sound is the target sound. Yet another suggestion, basically being an exact auditory representation of the arrow, was to include short directional beeps at close ranges, being activated automatically (e.g. when there is little time left). We concluded that it is better to include all in-game sounds in the actual game (i.e. not the VoiceOver suggestion), and that the second suggestion was most likely the best, as it would give a more natural feel and could be used more extensively, whereas the directional beeps could also be useful, but that they would not give as much of a distinction between front and back. However, as both methods might work (either under different circumstances, or perhaps in combination), they should therefore be tested before drawing definite conclusions.

**5.7.3.3 Perception**

**Distance:** P6 and some of the other participants in the usability tests commented that it was easier to hear the direction of the sound source at closer ranges. This is completely natural, as it becomes more difficult to perceive the various cues the further away the sound is (e.g. when hearing a bee at a distance of 8 meters, it will be more difficult to determine its direction compared to a distance of 1 meter, simply because the sound level has been halved three times, thus making it difficult for us to perceive cues such as IID’s and spectral differences). Therefore, I asked P6 whether he thought the maximum distance in the game (8 m) was too long, which he did not think. He also thought that the meta-level communication as used in Sound Hunter is very effective for increasing the perception of distance. However, we also discussed other possible ways of making the perception of direction to sounds at further distances more audible, such as making the sound move around, as opposed to being static. In this case, smaller movements would make it easier to perceive direction, whereas circular movements around the player would make navigation more difficult, which could be used as a means to increase the difficulty level.
**Increase audibility:** During the usability tests, P3 had difficulties hearing level 20, and P2 could not hear level 20 at all (and barely perceive level 19). When discussing these levels, we came to the conclusion that they could be kept, but that the frequency ranges have to be much tighter. There are at least three good reasons for this; to eliminate differences in difficulty between ages, as this imposed difficulty is unfair to the player, that most commonly used earphones do not have a frequency range of 20 to 20kHz, as well as to get rid of the noise imposed by the earplug–filter occurring at very high frequencies (around and above 15kHz).

**Silence:** All the participants in the usability tests thought that short sections of silence in the audio loops made the game more difficult, where some participants felt that this made them focus more on their hearing, and others made use of the silence when playing the game again, as they knew how long to wait. However, even though the levels were sorted according to their difficulty, there might be a better way to make use of silence. We concluded that silence might be used to increase the difficulty in the game, but that care has to be taken in order not to cross the border of confusion, and that the amount of silence in the loops should increase slowly, but always remain under five seconds. Furthermore, we concluded that silence, in general, should never be used such that navigational aids (e.g. directional beeps mentioned above) are required, as this would be equal to removing an introduced problem by including more sounds, in which case there are other ways to make the game more difficult.

**Clicks and noises:** During the focus groups and usability tests, I wanted to see how many participants noticed noise or clicks imposed by the filter on levels including very high frequency content. Few participants noticed imposed noise at all, but among the three who did, two of them (one being P6) thought that this was negative for the game. We therefore came to the conclusion that these levels should be either low-pass filtered to avoid the highest frequency components, or be completely removed – “These are artefacts and should be removed, as they expose a problem in the code, which is not so good”.

**Tiredness:** All of the participants except P6 felt that they focused more on their hearing and that they trained their hearing (answers ranging from “to some extent” to “very much” on a 7 point likert scale). On the positive side, Sound Hunter may be used to focus more, or train a person’s hearing. However, many participants also stated that they became tired (usually after about 40 minutes). One of the sighted players suggested that including more levels might counteract this. When asking P6, he stated that – “Definitely. When I played the game the first time, I thought it was very fun, but it became repetitive when playing five sessions in a row. Any game or task like this requires concentration, and 40 minutes is usually the amount of time you can spend on such a task before you become tired, or start thinking about other things, such as with lectures for example”. We therefore concluded that not only would more levels be beneficial for Sound Hunter, but also that there should also be a “Pause” button. Depending on the amount of levels, the game could also include various “saving points”, where the player can begin when dying.

5.7.3.4 **Ideas for future development**

**Introducing elevation:** We discussed how elevation might be used in Sound Hunter (e.g. having a sound move up and down, forming a wave when turning the head). However, after testing the elevation, we came to the conclusion that it was quite difficult to perceive (generally, sounds above and below are more difficult to perceive than sounds around the listener), and that if using the elevation, sounds should not be placed at, or moved quickly between extreme angles (e.g. 90 degrees above or -40 degrees below), but rather move carefully between above and below 0 degrees (e.g. -20 to 20), as this is more likely to give a better perception of elevation. Again, however, this is something that requires further testing.

**Split the game into different “Sound hunting missions”:** This idea had to do with including more levels and categorizing them in different ways related to making the game more difficult, such as the amount of realness or silence in audio loops, perceptibility in pure tones, complexity in meta-level communication, or room size with the corresponding distance and decay-time for beeps and clicks. P6 thought this was a great idea – “All of that sounds great. It’s a classic way of getting people to play more so it’s a really good idea”. We concluded that the best way of doing this would be to include the sound hunting missions in the game menu (i.e. when pressing
“Play”, the player is presented with “Sound hunting mission 1”, “Sound hunting mission 2” and so on), and that the later levels would be locked until the first level was completed. Also, in order to encourage the player to perform better, the “unlockables” would be linked to performance, where the player would be able to unlock two missions when getting a perfect result on the first mission. P6 also stated that having a cave could be fun – “I would also like to have a cave, where an extra component could be to have walls. When hitting a wall, a sound is heard, meaning that the player has to navigate through a maze and capture sounds etc.”.

**Introducing more sound objects in the levels:** This was one of my original ideas, which unfortunately I never had the time to implement, but it would definitely be an interesting addition to the game. The basic idea is that the player hears many sounds at once, but has to capture them in the right order (e.g. by having one “goal sound” that is clearly distinguishable from the other sounds, where all the non-goal sounds may be similar in sound character, and have to be captured before the goal sound, which could be done by having easier levels with only the goal-sound, the goal sound + 1 non-goal sound, the goal sound + 2 non-goal sounds, and so on). P6 also thought this was a good idea, and added a very interesting component – “You could have a chord or a song that the player has to capture, bit by bit, and if the wrong bit is captured, the player dies. This is a bit like the game “Jungle George”, but as you are working with 3D-audio it would be even more fun”.

**Artificial intelligence:** Several participants in the usability tests suggested that it would be fun to make some kind of car game or airplane game. This could be done by introducing more sound objects (as mentioned above) where the player receives points for catching certain sound objects, while avoiding other threatening sound objects, either moving around in the same space, or also shooting at the player. P6 also came with a further suggestion, involving the elevation – “The player could also be a mouse on the ground, where the mission is to capture small insects, while avoiding the bird flying around in the air, occasionally diving down to hunt the player”.

Even though these suggestions sound fun in theory, they have to be carefully developed and tested. The more sounds are involved, the more consideration has to be taken as to how these sound objects behave, how the player distinguishes threatening sound objects from those to be captured (e.g. as mentioned above, by introducing one at a time), as well as maintaining the possibility to efficiently navigate solely by using 3D audio. Therefore, the main conclusion from this session was that all of the more advanced suggestions for further development and future 3D-audio games similar to Sound Hunter have to be developed such that they follow the principles of the suggested framework.
6 Conclusions

When playing an action-based game, whether it is graphical or auditory, the player wants to be transported into another world and perceive that world as if it were real. I argue that an increased level of immersion due to increased imagination is not necessarily the reason behind the success of action-based graphical video games, and there is little reason why this should be the case for audio games. Creating a sense of auditory spatial presence can only be accomplished properly by connecting our real-world perceptions with the actions and responses in that of the game. Today, audio games based on 3D-audio are extremely rare compared to audio-visual games, with only a handful being available for iPhone users. For these games, the main problems seem to be the inability to use 3D-audio as the only means of navigation, reliance on stationary equipment or physical movement by the player, as well as insufficient usage of HCI methods in the development process.

In this thesis, I have presented a framework for designing 3D-audio games in which all navigation is based on perceiving the 3D-audio, as opposed to relying on other navigational aids or imagining the audio as being spatial, where additional sounds are added later on in the development process. In this proposed rethink of the design process of 3D-audio games, the focus is shifted from creating complex auditory environments in which 3D audio is used as a complementary spatial effect, to creating a functional 3D-audio navigation being complemented by carefully enriching the auditory environment. By following this design principle, audio-only games may be made to respond more rapidly and accurately to the player’s navigational input, as well as give the player a sense of full control and feeling of immediate response from the auditory environment, which could help bridge the gap between the hasty action-filled video games and the slower tempo audio games we have today. To test the framework, the game named Sound Hunter was developed in an iterative process through focus groups, usability tests, and a final expert evaluation, together with both sighted and visually impaired participants. A summary of the results can be found in the section below, followed by proposed areas for future research.

6.1 Summarizing the results

- Through the development and testing of Sound Hunter, I have successfully showed that it is possible to build a navigational HRTF-based audio-only game where all navigation is based on perceiving 3D-audio cues, as opposed to using additional sounds to aid navigation, thus answering research question 1.
- Even though the game was built as a prototype (i.e. a computer was necessary to run the game), it was built in such a way that it could easily be transformed into an actual iPhone application. The computer was not in any way part of the tests, other than to start the game, gather level times, and provide an audio output, all of which could be done by the iPhone (i.e. the game did not rely on head trackers or other stationary equipment in the same way as other navigational 3D-audio games have done so far). Also, the players could effectively play the game in a relaxed manner, without involving physical movement other than utilizing the iPhone’s accelerometer through simple hand movements. Therefore, I regard research questions 1.1 and 1.2 successfully answered.
- The accuracy of the participants’ navigations was measured as the mean radius percentage needed in order to successfully capture the sound while navigating without a time limit. Even though participants differed in their navigation abilities, all of them managed to capture the sounds within the set score radius of 5% of the total radius, equalling 0.4 meters. However, as many participants felt that it was easy to capture the sounds, they might very well have captured sounds at an even lower radius percentage (such as in focus group 2, where the participants managed to capture the sound within 2% of the total radius, or 0.16 meters). As this kind of accuracy is more than enough to prove that people can navigate...
with a high level of accuracy during interactive gaming experiences, I therefore regard research question 1.3 successfully answered.

- During the usability tests, the amount of tips, the quantitative data analysis, the observations, and the qualitative feedback suggested that the learning curve for the navigation was equal to one gaming session or less. Some participants immediately navigated perfectly, whereas the learning curve was slightly longer for others. When adding the practicing time (varying between participants, but with a maximum of about 5 minutes until they felt ready), the total time for learning the navigation varied from a couple of seconds to a maximum of about 15 minutes, thus answering research question 1.4.

- To answer research questions 1.5 and 1.6, there are three main differences between stereo and HRTF synthesis being relevant for audio game development, where the first is that of localization. With stereo, localization in all angles is impossible, as the sounds can be panned only between left and right. This is referred to as lateralization, and means that sounds are perceived as being positioned somewhere on a straight line between the ears (i.e. they are internalized, or “inside the head”). HRTF synthesis can be seen as a way of digitally representing how our natural hearing works, by filtering the sounds through convolution such that various navigational cues caused by the pinnae, head and upper body alter the characteristics of the sound, which in turn makes it possible for us to perceive the sound as coming from a certain distance and direction in space (i.e. they are externalized, or “outside the head”). As the technique is based on separating the signals to each ear from one another, headphones or earphones are necessary in order to properly perceive the effect (as opposed to stereo where speakers may be used), which is also the second main difference between the techniques. As perceiving sound objects as coming from a certain place is usually not the main priority when using stereo, more sounds can be used simultaneously (e.g. background music, attacking enemies, or voices). However, when using HRTF synthesis, additional sounds may severely disturb the ability to localize other sounds when visual cues cannot be used. This is also the main problem with the way in which audio-only games today are built, where it is common to include over 30 sound tracks being played back simultaneously, even when making use of HRTF synthesis. The key misconception among audio game developers is therefore that HRTF synthesis may be added as a complementary spatial effect in order to enrich the auditory environment, to make it sound more realistic. However, there is a third main difference between stereo and HRTF synthesis making this difficult, which is the fact that either player head movements, or moving the sound source, is necessary in order to properly perceive the 3D effect. As opposed to stereo, HRTF synthesis is therefore linked to the navigation, meaning that all player movement has to be connected to each and every sound, if the intention is to provide the player with the perception of a truly 3-dimensional spatial environment, which is also the exact opposite of what commercial 3D-based audio-only games today do. In order to truly make use of HRTF synthesis for the purpose of navigation and true 3D-audio perception, games have to be created on the foundation of the three above-mentioned principles. Instead of adding 3D audio as a complementary spatial effect, a functioning 3D-audio navigation has to be built, followed by carefully enriching the auditory environment. If several spatial sounds are to be played simultaneously, they all have to be HRTF filtered and connected to the navigation. Depending on the circumstances, it may be possible to mix stereo sounds and HRTF-filtered sounds, such as having ambient background sounds of a forest in stereo at low volume while HRTF filtering a more dominant sound. At the same time, this seems unnecessary, as if the goal is to be able to navigate in the auditory environment, the ambient sounds should also be filtered. Therefore, when wanting to add more spatial sounds, the navigation must always be usability tested to confirm that it still works, or else the 3D-effect will most likely be ruined, or be of lower quality. However, this does not mean that stereo sounds cannot be used at all. As in Sound Hunter, stereo sounds may be used for everything not having to do with navigation, such as auditory icons or earcons to convey confirmatory sounds related to player actions, or sonification to map abstract data into static or dynamic auditory communications.
• Understanding the difference between HRTF synthesis and stereo, making use of reoccurring iterations of testing and evaluation, as well as acquiring a fundamental understanding of our human sound perception are therefore all extremely important when developing audio-only games. This was also the foundation for the creation of the framework used to develop Sound Hunter. Furthermore, as the clear perception and possibility to navigate in the 3D-audio environment in Sound Hunter was extremely appreciated, even by an expert in the field of audio game development, who claimed that it was the best 3D audio he had ever experienced in a game, I consider the framework a successful modus operandi for similar development projects, and there is little reason to doubt that it should also be applicable for general audio-game development purposes. Therefore, I regard research question 1.8 successfully answered.

• To answer research question 1.7, the qualities of the sounds used in Sound Hunter to make the game easier or more difficult, were; the sounds connection to real-life scenarios (where a stronger connection was considered easier, and a loose connection was more difficult), the level of temporal variability in the sound (where big changes to the sound over time was considered easier, and small changes more difficult), the amount of meta-level communication in the sound (where a complex communication was considered easier, and a less complex communication more difficult), the amount of silence in the sound (where constant playback was considered easier than playback every now and then), and finally the frequency characteristics (where a wide range and audible frequencies were considered easier, and closer ranges as well as frequencies bordering on audibility difficulties were considered more difficult). The results showed that all of these sound characteristics might be used without ruining the 3D-effect, but that consideration has to be taken as to how these characteristics are altered. For example, even though the results showed that meta-level communication, as used in Sound Hunter, was very effective for increasing the perception of distance and direction, careful consideration has to be taken as to how it should be varied. If using meta-level communication in the wrong way, this might lead to illusions, making navigation impossible. This is a very important conclusion when it comes to designing 3D-based audio-only games, as it is desirable to increase the difficulty, but at the same time very undesirable to increase confusion and illusions. The illusions of false distance typically occur when sounds vary such that they became less perceptible the closer they come, and alterations of this type should therefore never be used in 3D-based audio-only games. Instead, other ways of communicating distance changes can be used, such as to include small movements in the sound object, or using appropriate reverberation in different ways, which could be done by linking reverberation to meta-level communication, by increasing the decay for further distances. Off course, doing this would not be realistic, but on the other hand, none of the meta-level communications used in Sound Hunter were realistic, yet still very much enjoyed by the participants. Furthermore, the results showed that a higher temporal variability in the sound made it easier to localize, where all the sounds having a higher oscillation were easier to localize than those with a low oscillation. Regarding silence, the participants had mixed feelings about it, where one thought it was negative, some thought it made them focus more on their hearing, and others used it as a form of clue during later gaming sessions. All participants regarded silence as being difficult however, which in the case of Sound Hunter is quite natural, seeing as there was no clear progression, and the only aid through the silence was to follow the various reverberations and echo decays. However, silence can be used in audio-only games, but in this case it should increase progressively such that the border of confusion is not reached, which appears around five seconds of silence. Furthermore, additional sounds should not be used to cover up or aid navigation through silence, as the occurrence of this need implies that silence has become a problem. Finally, the frequency of the sounds were also successful in making the levels more difficult, where wider ranges made levels easier, as well as frequencies being more audible. When it comes to the audibility of frequencies, making use of very low or very high frequencies in audio games works, but careful consideration has to be taken as to the character of these frequencies. Using the most extreme frequencies (e.g. bordering 20
Hz or 20 kHz) is only a good idea if the frequencies vary in a way such that they often become perceptible. Otherwise, people in different age groups, or those using earphones or headphones with a narrow frequency range, will not be able to perceive the sounds. If using the [earplug~] external in Pd, sounds including very high frequencies (above 10 kHz) should be either low-pass filtered, or not be used at all, as they lead to external noise being imposed by the filter. Furthermore, altering between easily perceptible frequencies and frequencies being more difficult to perceive is likely to lead to negative distance illusions having been mentioned, and this should therefore never be done in a way other than how levels 19 and 20 in Sound Hunter were constructed (i.e. even if positioning a sound at a fixed distance and varying between perceptible and extreme frequencies, the same illusions will occur, and therefore, the frequencies should always be altered such that they become more easily perceptible the closer they come).

- The results of this thesis indicate that visually impaired people are interested in audio games, and that the amount of audio-games available for them is very limited. Also, almost all audio-only games intended for visually impaired people today are developed for the PC, while the number of visually impaired smartphone users is heavily increasing. Therefore, there is not only a strong need for audio-only games intended for visually impaired smartphone users, but also a great market potential for audio-only games directed to visually impaired iOS users, answering research question 2.

- When developing iOS games for visually impaired people, one has to realize that with the help of modern technology, such as the VoiceOver function, visually impaired people can navigate through menus in the same way sighted people can. The only real difference is that it takes slightly longer (i.e. scanning through every menu item one at a time, as opposed to getting an overview of everything at once), as well as that pictures and images cannot be represented (i.e. the VoiceOver function will simply say “image x”). Therefore, when creating game menus, these may be created similarly as for sighted people, with the exception that describing images should be excluded, and all the relevant information should be communicated in button labels, or through text. This also means that including pre-recorded voices for instructions or other menu-related information is unnecessary. Furthermore, it is always good to make the menus as efficient as possible, for example by not including too many buttons and options, as well as including a well established default setting for the controls, which may be adjusted, but with as few parameters as possible (i.e. developing a good standard is better than including total customization). If wanting to utilize the iPhone in different angles (e.g. horizontal instead of vertical), this is no problem, as the VoiceOver informs the user about the iPhone’s holding position. Therefore, when developing games being based on the iPhone’s accelerometer data, this is not a problem. Even though visually impaired iPhone users are more used to holding the iPhone more upright in a vertical position, they easily adapt to other holding positions, as long as this holding position is the optimal choice for utilizing the accelerometer data in the game. However, the most crucial understanding related to the developing of iOS games intended for the visually impaired, is that usability testing has to be included throughout the process in order to ensure that the controls, or other aspects related to the game are optimized according to their preferences, answering research question 2.1.

- Despite the simple game character of Sound Hunter, all of the participants felt happy while playing the game, and were either highly immersed in the game, or in a state of flow. This indicates that 3D-audio, as provided in a simple game, still has the ability to provide a high level of immersion. This is very promising for 3D-based audio-only iOS games, especially as most of the participants did not even play the finished game, where the levels were in the correct order, the time limit was included, and further auditory icons were activated. Furthermore, all of the participants enjoyed the game and would have recommended it to others, and both sighted and blind participants were very impressed with the 3D-audio experience. For the sighted participants, the concept of playing an immersive 3D-based audio-only game was highly appreciated, mainly as it was seen as something new and exciting. For the blind participants not playing audio games, the experience of simply being
able to play a motion-based game as a blind person was seen as something very positive and fun, whereas the blind participants used to audio-only games were positively surprised and impressed by the accuracy of the 3D-audio rendering, and how it could be used to navigate to capture sounds. These results suggest that HRTF synthesis, when used in the right way, most definitely adds something lacking in today’s audio-only games, thus answering research questions 2.2 and 2.3.

- When it comes to how the game may be developed further in order to become as entertaining as possible, the results indicate that this mainly had to do with the characteristics of the game. As mentioned, the 3D-audio experience was highly appreciated by all participants, but at the time of testing, the game in itself was not fully developed (i.e. making it a proper challenge in time, or providing the ability to beat the current highscore), and it was therefore more a matter of capturing a series of sounds without any real imposed challenges except locating the sound object. My initial intention, as well as the final version of the game, was for it to be simple arcade-based game, where the player captures sounds while constantly trying to beat their time and highscore record. For visual iOS games, these types of games exist in large quantities, and many of them have become very successful, despite their simple character. Also, as so many of the participants were highly immersed in the game even when only capturing random sounds with no time constraint, it seems as if there is great potential for simpler 3D-based audio-only games. After all, smartphone games tend to be of simpler character, and that is what most people expect. Of course, this mainly applies to sighted people, as a real standard is yet to be set for visually impaired smartphone gamers. However, there is definitely room for improvement when it comes to the purpose of the game (i.e. “why must I capture these sounds?”), where simple alterations could make a big difference in the enjoyment of the game. For example, some participants suggested that car or airplane games would be great for the type of accelerometer-based navigation that is used in Sound Hunter, and games could therefore be made such that the player drives a virtual vehicle, where the smartphone acts as steering wheel. Furthermore, adding only a few sound objects could make the game more interesting. For example by imposing more challenge, where the player has to avoid certain sound objects while capturing the target sound object, or by changing the goal of the game, where the player has to capture several sound objects in the correct order (e.g. building a chord or song), or even just the simple change of having one looping sound object alternating in its character, where the player only is able to capture the sound object when it sounds in a certain way (e.g. crying baby vs. happy baby). Also, as the levels varied in difficulty, a good future development of Sound Hunter would be to include different un-lockable sound hunting missions, which could be created by varying the amount of realness or silence in audio loops, perceptibility in pure tones, complexity in meta-level communication, or room size with corresponding distances and decay-times for beeps and clicks. Finally, easier forms of artificial intelligence could be used, where threatening objects attack or shoot at the player. Of course, all suggestions involving more sounds simultaneously may lead to navigation difficulties, and therefore, it is recommended that all of these improvements follow the principles of the design framework presented in this thesis, such that the navigation is still functional, thus answering research question 2.4.

- To answer research question 3, the results of this thesis support that HRTF-based audio-only games may be used to train people’s hearing in an entertaining way. All the participants except one, even those who were blind, felt that they focused more on their hearing and even trained their hearing while playing the game. The only problem was that participants became tired after 30-40 minutes of gaming. However, this is natural, and can be easily compensated by including a “Pause”-button or “Save”-button depending on the amount of levels.

- Even though it was seldom the case in this thesis, it is common for people to experience front/back-confusion during passive azimuth localization, which can be explained by the fact that they cannot make use of head movements or visual cues, whereas for interactive experiences in daily life, it has been shown that people immediately adapt to the azimuth
when wearing foreign HRTF implants. However, no studies have shown how long it takes for people to adapt to a foreign HRTF under virtual interactive experiences (e.g. when used for navigating in games), and there is reason to believe that this should take longer, seeing as the virtual environment will not provide navigational cues being as accurate as in the real world. In this thesis, the participants’ ability to adapt to a foreign HRTF while playing Sound Hunter was tested, where the results indicate that the difficulty of each level and adjusting to the navigation, were the main factors affecting the participants’ abilities to adapt to the foreign HRTF while gaming. Even though participants had different preferences for the controls, they all used the navigation to eliminate front/back-confusion through spinning the sound. This was easier for those who understood the controls, who seemed to have no problem with front/back-confusion. For other participants, it took slightly longer. However, all the participants navigated better and perceived front/back better when playing the easier synthesized levels, where the difference was small for some participants, but very big for other participants. The reason why the easier synthesized levels made it easier to perceive front from back, was because the complexity of the sound changed depending on its distance, giving more clues related to the position and distance to the sound object. The results therefore indicate that the main factor affecting the learning curve for navigation, is to optimize the navigational controls to the player’s preference, and the main factor affecting the learning curve for azimuth localization in virtual interactive experiences, is to find ways of eliminating front/back confusion. However, the exact relationship between these two learning curves is more difficult to exactly predict from the results. Even if one of these learning curves is optimized, this may not lead to the other automatically being optimized as well (i.e. even if the controls are optimized to the player’s preference, the player may still have problems with front/back confusion). Therefore, the best way of solving this problem is to optimize both of them. The controls should be made as intuitive as possible and include a good default setting, where the user may customize certain parameters to their personal preference (e.g. including a sensitivity slider as in the case of Sound Hunter). When it comes to preventing front/back-confusion, the results showed that complex meta-level communication is very effective. However, there might be many more ways of preventing front/back-confusion, for example by including front/back sound bounces, or directional beeps. Even though these methods were never tested, they were discussed in the final evaluation, and seem promising for future development, thus answering research questions 3.1 and 3.1.1.

- As the participants found the sounds to be very or extremely externalized, the results indicate that the HRTF-filtering was successful in providing an out-of-head experience, even for synthesized sounds. This is very promising for future audio-only games intended to train a person’s hearing, as it implies that HRTF synthesis may be used to provide more realistic virtual scenarios, and to train the full scope of a person’s hearing (i.e. proper localization added), in a fun and interactive way, thus answering research question 3.2.

- Even though some of the participants felt that the sound file levels sounded more realistic, the participants also found the synthesized sound levels equally entertaining as the sound file levels. Generally, when developing audio-only games also being useful for training a person’s hearing, the focus is not to make the game sound like a hearing test, but simply to provide an entertaining game, allowing the player to focus more on their hearing by their own will. However, if synthesized sounds can be used such that they are equally entertaining as sound files, which the results suggest, this can be very useful, as synthesized sounds allow full experimental control over every parameter in the sound, thus making it possible to separately train and monitor the results for every aspect of the sound being relevant for proper recovery, such as singling out a certain frequency, or performing various masking tests, audibility tests, or other forms of perception tests in virtual environments, with or without the constraints of real-world physics, thus answering research question 3.3.

- Finally, the results indicate that people with ADHD might have difficulties in playing HRTF-based audio-only games requiring concentration. However, the results also indicate that if the levels are designed correctly using meta-level communication, these difficulties
are largely reduced, as long as the sound captures and continuously maintains the person’s concentration, where the slightest deviation from this continuous change makes the game much more difficult, and even seems to raise the absolute threshold for binaural sound perception. This is therefore an area that could be greatly aided by future research, as covered below.

### 6.2 Future research

During Sound Hunter’s development process, I have come across a variety of areas that would be interesting to conduct future research in. In this section, I will therefore summarize these areas.

**Spatial perception among people with ADHD:** As mentioned above, people with ADHD seem to have greater difficulties in perceiving spatial sounds, as well as navigating in HRTF-based virtual environments, as long as the sounds are not continuously changing with appropriate meta-level communication. As it is common for visually impaired people to also have other forms of disorders, such as ADHD, further research within this area including a larger number of participants, could possibly reveal important information on how audio-only games can be designed best for people with ADHD. It is also possible that audio-only games have the potential to progressively train people with ADHD to focus more concentration to what they hear, for example by slowly changing the characteristics of the meta-level communication in order to find the border at which concentration is lost, followed by trying to raise this border.

**How to make use of reverberation and echoes:** All the participants achieved better results for the easier synthesized levels, but even though they might have been easier to localize, some participants (mainly the visually impaired) thought that they were less externalized than levels that sounded more real. This indicates that there is a correlation between a sound’s realness and the perception of externalization. The [earplug~] HRTF-filter seems to be very efficient for localizing sounds in 3D, with the exception of sounds including very high frequency content, or sounds having a bad initial sound quality (e.g. low bit-depth), which causes noises and clicks in the filter. The reason why the [earplug~] filter is so effective, however, is difficult to say. One reason could be the lacking representation of a real acoustic space (i.e. room size with proper occlusions, diffractions and reverberation), and it may very well be the case that these kinds of simulations are not appropriate for audio-only games where the focus is on 3D-audio navigation, simply because they lead to too many sounds being present in the mix, thus making it difficult for the listener to perceive distance and direction. As it is difficult to find out exactly how the more advanced systems make use of HRTF synthesis (e.g. if they filter all sounds including reflections, or if they apply room effects afterwards), it becomes difficult to know what the differences are between these systems and [earplug~]. In the case of Sound Hunter, the levels including perceivable echo effects (e.g. spooky water drops and pterodactyl screech) were filtered with all echoes included. As all frequencies and reverberations are not directional in real life, this might not give a correct representation of room effects. For example, it might be better to apply some of the low intensity reverberation separately (especially the lower non-directional frequencies), as it might give a more dynamic auditory landscape (e.g. standing in a church, where some sounds are directional and some lie in the background to give the feeling of a big echoic chamber, where low intensity sounds are heard in stereo, or from all directions). On the other hand, the separately added reverberation cannot be too strong, as it would make it more difficult to localize sounds. The relationship between realness and externalization, and the best way to make use of reverberations and echoes without ruining the 3D-audio navigation should therefore be studied further, where including visually impaired people would aid the process.

**Further tests with greater sample sizes:** If games similar to Sound Hunter are made in the near future (or if I have the possibility of releasing Sound Hunter as an iPhone application), similar tests as those in the quantitative data analysis could be performed, but with greater sample sizes. Using larger sample sizes would be valuable in order to find significant results for the
differences between sessions, the relationship between game improvement and level order, or which levels are the most efficient for training people’s hearing. A large quantity of participants could also be used to evaluate the game in questionnaires, to generalize findings over a greater population, as well as to rule out various correlational variables.

**Further ways of utilizing the accelerometer function:** During the development of Sound Hunter I examined two different ways of utilizing the accelerometer function together with the user group, where easily understood hand movements and customizable sensitivity seem to be the most important aspects for the controls. However, there may be many more, better, ways of utilizing the accelerometer function. For example, if the intention is to mimic our usage of head movements when localizing sounds in real life by using accelerometer data, this could be done by letting the player hold the iPhone against the side of the head, where alterations in accelerometer data caused by physical head movements could be wired to the azimuth and elevation. The results from this thesis indicate that this might be difficult however, mainly due to the long period of time required to adapt to the elevation, but also because players, especially the blind, tend to prefer simple controls not requiring extensive physical movement. Promoting head movement could therefore become problematic. Still, future research could examine the possibility of utilizing the accelerometer’s z-data (or a combination of the accelerometer data and the gyroscope data), to promote a more intuitive navigation.

**Increasing perception of elevation in virtual interactive environments:** Adapting to the elevation of a foreign HRTF in the real world takes a long time (i.e. several weeks), and it could therefore be argued that elevation, in fact, cannot be utilized for the purpose of audio-only games. However, it might be possible to slowly train a person’s perception of elevation in a 3D-based audio-only game or virtual interactive environment, by using smaller and more easily perceptible alterations at first (e.g. -20 to 20 degrees or less), followed by slowly increasing these angles in a way matching that individual’s perception level (for example by letting them adjust the elevation range themselves). The sounds could be either positioned, or slowly moved up and down, thus creating a wave when the player turns their head. A more futuristic alternative to this would be to create the possibility of utilizing the player’s individual HRTF. In the same way modern 3D-printing technology has evolved, the possibility of letting the players scan their own heads could be offered (or simpler methods, such as taking facial pictures with the computer camera and using face/head recognition to build a head model). This head model would then be linked to a database including millions of impulse responses, where those matching the player’s head the best would be picked. If this would be made possible, 3D-based audio-only games could become truly 3-dimensional, where complete auditory environments could be represented, or sound objects could be diving down towards the player.

**Eliminating front/back confusion:** As mentioned in the report, additional ways of reducing front/back confusion is always good (e.g. including additional sounds, such as front/back sound bounces or directional beeps), and future development and testing of various ways of doing this would greatly aid navigation in HRTF-based 3D-audio games.

**VoiceOver in audio-only games:** Furthermore, how the VoiceOver function works when being implemented in an audio-only game should be examined, as if it works, this could lead to less auditory icons having to be used, as well as being able to include a range of different languages in the same game (e.g. walking up to a person followed by holding the finger on the iPhone screen to hear what he has to say, being read by the VoiceOver in English or Spanish).

**Localization and memories of visual cues:** Visual cues greatly aid localization, but according to the results in this thesis, it seems as if even memories of visual cues may aid localization, as one of the participants expressed that he tried to “see the sounds”. Of course, even if memories of visual cues in fact do increase the ability to localize sounds, this is difficult to prove, since even if a study would be conducted where people with born blindness were compared with blind people having memories of visual cues, it could be argued that people with born blindness have better localization abilities as they have trained their hearing since birth, and former sighted people have better localization abilities as they have memories of visual cues. Nonetheless, it is still an interesting finding, as it illustrates the conflicts between these two groups of research.
6.3 Last words

Sound Hunter only scratched the surface of what may become the future era of fully navigational 3D-based audio-only games, relying on perception rather than imagination. The development of this game and the research I have conducted has been one of the most interesting and rewarding experiences of my life, and I hope those who have read either this thesis, or the scientific article having been accepted and published for the joint SMAC/SMC conference 2013 (Brieger 2013), have found it equally interesting. For further details on the programming of the game, or other questions, please contact me on brieger@kth.se.
7 REFERENCES

Abeyasekera, S 2000, 'Quantitative Analysis Approaches to Qualitative Data: Why, When and How', Statistical Services Centre, University of Reading, United Kingdom.


Bell, J 2009, Introduktion till Forskningsmetodik, 4th edn, Studentlitteratur AB.


Röber, N & Masuch, M, 'Playing audio-only games a compendium of interacting with virtual auditory worlds', Department of Simulation and Graphics, University of Magdeburg, Magdeburg, Germany.


Appendix (Swedish)

Fokus grupp 1

Sätt igång mikrofonen, ta fram anteckningspenna/papper
a. Välkomna! Tack för att ni är här, jag tror detta kommer bli riktigt kul!
b. Tack Funka för lokalen och initiativet!
c. ”Jag har aldrig umgåtts med blinda, så säg gärna till om jag är överdrivet tydlig, eller om ni vill att jag tydliggör någonting”

Hälsa alla välkomna, presenterar mig själv, ni presenterar er själva

Vem är jag?

d. Sebastian, KTH, ljud, MDI, musik
e. Under min tid på KTH:
   i. Arbetat mycket med 3D-ljud (förklaras efter deras pres.)
      1. HRTF i datorspel
      2. Mixa musik i 3D
f. Varför vi är här idag:
   i. Vill skapa ett ljudspel i 3D (förklaras efter deras pres.)

Gå runt i cirkel, låta er presentera er själva (namn och vad som intresserade er för att komma hit idag)

g. ANTECKNA: Rita en cirkel och skriv deras namn, intressen

Vad är 3D ljud?

   HRTF – Simulera människans riktningshörsel
   h. 2 sätt, docka, syntetiserad HRTF
   FRÅGA: Hur många här har hört någon form av 3D-ljud?
      i. (ex. Virtual barbershop)
   j. ANTECKNA/SÅG: Antal handuppräckningar
   k. Ev. FRÅGA: Var det i hörarlurar? Hur upplevde ni att det fungerade?

Spelet – Vad är det jag har tänkt?

   Spelindustrin, fokuserad på visuella spel
   Vill införa en ny dimension till ljudspel med 3D-ljud
   Vad kan 3D-ljud tillföra ljudspel?
      l. – Mer spänning/immersion
   m. – Skapa arkadspel, actionspel, äventyrsspel
   n. – Spel där man springer runt, flyttar ljudobjekt etc.
      i. FRÅGA: Vilka av er har spelat ett ljudspel någon gång?
      ii. ANTECKNA/SÅG: Antal handuppräckningar
      iii. GÅ IGENOM PTP: Vilka spel har du spelat och vad upplevde du var extra kul med dem?

   Plattform/spelkontroll = Smartphone
   o. FRÅGA: Hur många av er har en smartphone?
   p. ANTECKNA: Antal handuppräckningar

Generella uppfattningar – Vad skulle 3D ljud kunna tillföra i ett ljudspel?

   GÅ IGENOM PTP: Vad tror du vore extra roligt att göra med 3D-ljud i ett ljudspel?
   Ev. Ni som har spelat ljudspel, vad tror ni hade varit bättre i ett favoritspel om det hade varit tredimensionellt?

Ljudspel för att träna hörseln

   Bra för personer som har haft dålig hörsel och plötsligt får bättre hörsel (hörapparater/cochlea-inplantat etc.)
Det man gör: Använder sig av olika ljud som tänjer lite på gränserna för vad en människa kan höra eller inte (lågfrekventa basljud/högfrekventa diskantljud etc.)

I spelet skulle man kunna använda detta för att öka svårighetsgraden

Vad tror ni om detta? (Tänkte bara kolla generellt)

q. HANDUPPR. DISKUSSION

Jag har blivit mer intresserad av denna aspekt (träna hörseln) sedan jag började med detta projekt

Specifika åsikter kring spelandén

Min idé: Simpelt men underhållande arkadspel
Ett ljudobjekt placeras i en 3D-rymd
Spelarens uppgift: ta sig till ljudobjektet så snabbt som möjligt

r. När man kommer tillräckligt nära ljudet → ”pling!”

s. Varje bana:
   i. Nytt ljud (svårare och svårare)
   ii. 20 sekunder på sig på varje bana

1. >20s → game over

   t. Ju snabbare man klarar banan, ju fler poäng får man
      i. Reflekteras i olika ljud

Luta iPhonen åt vänster/höger = luta huvudet åt vänster/höger
Luta iPhonen framåt/bakåt = gå framåt/bakåt i spelet
Både ljudfiler och syntetiska ljud används

u. Ljudfiler: Låter verkligare (ex. instrument, myggor, etc.)

v. Syntetiska ljud: Kan användas för att träna hörseln och göra spelet svårare

Varje gång man slår sitt rekord hör man ”highscore!” eller ”nytt rekord!”

SYNPUNKTER, FÖRSLAG, ANPASNING TILL TIDIGARE DISKUSSIONER, BRA/DÅLIG IDÉ? (HANDUPPRÄCKNING)

Tack!

Vilka av er skulle kunna tänka sig att ställa upp på fler fokusgrupper (2-3 till) och användartester av spelet?

Fokus grupp 2

Sätt igång mikrofonen, ta fram anteckningspapper/penna

Välkomna igen P1 och P2!
Denna gång tänkte jag att fokusgruppen ska vara tvådelad:

Första delen:

w. Vad vi kom fram till förra gången
x. Vad jag har gjort sen dess

Andra delen:

y. Lyssna på teknologin
z. Testa navigeringen med iPhonen
aa. (Om vi har tid) Prova att spela några banor

FÖRSTA DELEN

Kortfattat, vad vi gjorde förra gången

- Lärde känna varandra
- Jag förklarade hur människans ljudperception och tekniken jag använder fungerar
   o Ni sa båda två att ni hade hört tekniken vid något tillfälle via Pink Floyds skiva
- Generella spelidéer, vad man skulle kunna göra
   o P1: Avancerade ljudmiljöer, andra användningsområden som GPS etc.
   o P2/P1: Allting som är roligt/underhållande vore bra
• P1 visade hur blindanpassningen fungerar i iPhonen, och vi pratade lite om uppläsningshastigheter med mera
• Jag förklarade om ljudspel för att träna hörseln, och att det i spelet skulle kunna användas för att öka svårighetsgraden
  o P2/P1: Skulle kunna vara bra, och att det beror på hur man lägger upp det, att det t.ex. ska öka i svårighetsgrad men inte uppfattas som ett hörseltest
• Sist diskuterade vi min idé
  o Simpelt men underhållande arkadspel
  o Ett ljudobjekt placeras i en 3D-rymd
  o Spelarens uppgift: ta sig till ljudobjektet så snabbt som möjligt
    ▪ När man kommer tillräckligt nära ljudet → "pling!"
    ▪ Varje bana:
      ▪ Nytt ljud (svårare och svårare)
      ▪ 20 sekunder på sig på varje bana
        o >20s → game over
      ▪ Ju snabbare man klarar banan, ju fler poäng får man
        ▪ Reflekteras även i olika "poängljudd"
  o iPhonen som spelkontroll
    ▪ Luta vänster/höger = luta huvudet åt vänster/höger
    ▪ Luta framåt/bakåt = gå framåt/bakåt i spelet
  o Både ljudfiler och syntetiska ljud används
    ▪ Ljudfiler: Låter verkligare (ex. instrument, myggor, etc.)
    ▪ Syntetiska ljud: Kan användas för att träna hörseln och göra spelet svårare
  o Några viktiga saker vi kom fram till:
    ▪ P1: Jämföra sin highscore med andras highscore
    ▪ P2: Läsa upp spelinstruktionerna i olika hastigheter (exemplet med ljudböcker)
    ▪ P1 gav tipset om Applevis.com
      ▪ Jag kollade upp det, verkar bra, men hittade inget om 3D-ljud

Vad jag har gjort sen förra gången
  o Utvecklat navigeringen och banorna
    ▪ Testat olika ljud och metoder för hur man kan stegra svårighetsgraden
    ▪ Två olika huvudmetoder;
    ▪ Verkligt → overkligt:
      Gå från verklighetstrogna och igenkännbara ljud → ljud som inte låter verklighetstrogna eller inte är igenkännbara (kan inte kopplas till något verkligt scenario man känner igen ifrån det dagliga livet)
  o Exempel:
    ▪ Verkligt scenario: Man hör en gråtande bebis
    ▪ Semi-verkligt scenario: Man hör en morrande tiger
    ▪ Overkligt scenario: Man hör en jämrande zombie
  o Hörbara → ohörbara ljud
  o Knysar an till människans ljudperception, där man kan använda ljud som har:
    ▪ Olika temporal variabilitet: Stora kontra små förändringar i ljudet över tid (musik är ett bra exempel)
    ▪ Olika mängder tystnad: Spela ljudet konstant kontra då och då
    ▪ Olika frekvenser-span: Breda kontra nära frekvensspan, eller frekvenser som närmar sig gränserna för vad vi kan eller inte kan höra
    ▪ Olika intensitet: Ljud med hög eller låg ljudstyrka
    ▪ Olika maskeringstekniker: Införa andra ljud som blockerar ljudet man är ute efter, antingen på olika platser, eller att man exempelvis blockerar ett helt öra
    ▪ Man vill öka svårighetsgraden, men man måste akta sig för illusioner!
References

- Tänkt på hur man kan skapa highscore-systemet:
  - Tid + highscore för varje bana
  - Total highscore (nuvarande rekord)
  - Tidsrekord för att ha klarat hela spelet (eventuellt någonting man kan använda för att tävla med när man väl har klarat spelet, d.v.s. ”vem klarade spelet snabbast?”)
- Har tittat på olika sätt att visa text på iPhonen
  - Antingen skicka text till iPhonen, eller skapa ett interface med tre knappar (spela, instruktioner, highscore)
  - Ska prata med min handledare om detta när han återkommer från Japan
- Nackdelar med spelet i nuläget:
  - Problemet med fusk när man lutar telefonen åt båda hållen samtidigt, vilket också leder till en bugg att spelet börjar om från första banan
  - Vissa ljud (t.ex. ”game over”, ”new highscore!”) saknas

ANTRA DELEN

Testa HRTF perception
En person i taget, för att se om ni hör 3D-effekten (har inget att göra med er hörselkapacitet så känn ingen prestationsångest):
- Öppna LJUDSPELET.pd
- Ladda in ett trumset (MTD/wav loops/6/HH/), se att allt fungerar
- Förklara att tekniken funkar bäst när man håller huvudet rakt (alltså inte hukar sig framåt eller lutar sig bak, och att detta är någonting jag kommer ha i instruktionerna)
- Låt P1 resp. P2 lyssna i in-ear hörlurar och be dem peka i ljudets riktning 10 ggr
  - ANTECKNA: Antal rätt/fel
- Stäng ner patchen

2. Testa navigeringen
   Fri navigering
   - Öppna SOUND_HUNTER.pd
   - Tryck på den globala pd dsp 1 message boxen (om den inte redan är aktiverad)
   - Tryck på start sound and timer-bangen och låt dem leka runt med navigeringen så länge de vill
   - När de har provat klart, tryck på stop sound and timer-bangen
   - När båda är klara:
     - Fråga: Vad tyckte ni om navigeringen? (kunde ngt förbättras etc)

På tid
- Säg åt dem att nu ska de försöka klara banan så snabbt som möjligt (”säg till när du hör konfirmationsljudet”)
- Innan start, gå in i navigationen och bangla ljudet till en random position (gula knappen)
- Tryck sedan på start sound and timer-bangen och läs av tiden när de säger att de har klarat banan
  - ANTECKNA: Tid det tog för dem att klara banan så snabbt som möjligt

3. Testa banor
   - Tryck på den globala gröna start knappen när de är redo (testa så att det fungerar själv först)
   - 3 försök var att prova spelet, så pratar vi om det sen
   - Informera om fusk-buggen
   - Efter båda är klara:
     - Fråga: Lade ni lägger märke till någonting som ni tyckte var extra svårt med banorna?
4. Tack!
Fokusgrupp 3

Sätt igång mikrofonen, ta fram anteckningspapper/penna

b. Välkomna igen P1 och P2!

c. Denna gång kommer fokusgruppen vara tredelad, men störst fokus kommer ligga i praktiska aspekter:

i. Första delen:
   1. Vad vi kom fram till förra gången
   2. Vad jag har gjort sen dess

ii. Andra delen:
   1. Testa navigeringen igen
   2. Testa att spela spelet
   3. Testa lite olika banor i spelet

iii. Tredje delen:
   1. Idéer för fortsatta undersökningar

FÖRSTA DELEN

Kortfattat – Vad vi kom fram till förra gången och vad jag har gjort

1. iPhonen borde hållas på bredden istället
   a. Mer intuitiv spelkontroll
   b. P1 visade att iPhonens blindanpassning talar om vilken lutning den har → iPhone-applikationen bör starta så att den redan i menyn ligger i horisontalläge så att användaren förstår hur den ska hållas innan spelet/träningen börjar
   c. Fixat i prototypen (spelet), men jag har ingen applikation för tillfället

2. I menyn ska det även finnas en träna-knapp med oändligt lång tid + konfirmationsljud när man når målet
   a. **Fixat i prototypen (spelet), men inte i appen**

3. iPhonen ska reagera snabbare när man lutar
4. Ljudets rörelse ska accelerera ju mer man lutar
   a. 3-4 **fixat för när man snurrar ljudet** (iPhonen reagerar mycket snabbare och ljudet snurrar upp till tre gånger snabbare än ursprungligen när man lutar åt vänster/höger).
   b. **Jag valde att inte göra detta för framåt/bakåt-lutningen** (med motiveringen att det oftast är så i vanliga datospel att man går i en konstant hastighet men kan titta runt hur snabbt man vill).
      i. Detta är dock något jag vill att ni tänker på när ni provar nu så jag vet om det bör ändras (kan vara så att iPhonen bör reagera snabbare även vid framåt/bakåt-lutning, eller att det i ljudspel är bättre att kunna gå i olika hastigheter etc.)

5. P2: Ännu ett sätt att succesivt trappa upp svårighetsgraden; ha längre tid på sig på de första banorna och gradvis korta tiden per bana
   a. **Fixat:** På den sista har man 20s på sig och för varje bana innan den banan ökar tiden med 3s, så på första banan har man 1min och 17s på sig

6. Ljudet måste dö ut snabbare på samma radie för att simulera att det verkligen befinner sig längre bort
   a. **Fixat:** Kom på att sättet jag hade programmerat detta på sist gav en dålig illusion av avståndet till ljudet. Nu har ljudet lägre ljudstyrka vid längre avstånd

7. Området (radien) inom vilken man fångar ljudet är för snäv och borde ökas (så att man kan ta ljudet även om det befinner sig lite snett fram/bak och man går mot det)
   a. **Fixat:** Har gjort det lättare att ta ljudet, men vi får se vad ni tycker

8. Problemet med fusk måste lösas
   a. (Om man lutar framåt/bakåt och samtidigt åt höger/vänster ska ljudet till skillnad från att snurra runt snabbare och snabbare tills man får en perfekt
highscore fastna i en position, likt en osynlig vägg där ljudet inte kommer längre bort)

b. Fuskproblemet ledde även till en mängd andra problem, exempelvis errormeddelanden, förvirring om var ljudet befann sig, samt att andra Funktioner i spelet slogs ut (började om från första banan, konfirmationsljud även när man inte plockat ljudet etc.)

i. **Fuskproblemet löst:** Vid maximala avståndet stannar ljudet som om det krockar med en osynlig vägg, till skillnad från att dyka upp på motsatt håll).


9. Vissa ljud måste bytas ut då de har dålig ljudkvalitet (flugan och det högfrekventa ljudet av en dinosauriefågel)

a. **Bytte ut till en ny fluga och en ny dinosauriefågel i bättre kvalitet**

10. De sista banorna måste skapas (hade 18/20st varav 2 orsakade falska illusioner och behövde bytas ut, så 4 nya banor)

a. **Fixat:** 2 av banorna orsakade falska illusioner (kvarstod trots att jag fixade så att ljudstyrkan reglerades korrekt), så dessa togs bort och fyra nya banor lades till. Nu har spelet 20 banor (13 ljudbanor, 7 syntetiska)

11. Fixa ljuden ”game over”, ”new highscore”, ”new time record”, samt ”game completed”

a. **Fixa alla ljud, men har inte fixat så att ordningen är som den ska än** (exempelvis om man klarar spelet, får nytt tidsrekord och en ny highscore så måste de läsas upp efter varandra och inte samtidigt)

**ANDRA DELEN**

Testa övningsdelen igen

- Testkör så det fungerar
- Jag lyssnar samtidigt som de navigerar
  - Visa hur de ska hålla, placera ljudet
  - Ev. **Anteckna det som verkar vara lätt/svårt**

- Frågor om punkterna jag har förbättrat, samt generella synpunkter (**efter att båda har provat**):
  - ”Hur upplevde ni att navigeringen fungerade nu jämfört med förra gången rent generellt (bättre/sämre)?”
    - Är framåt/bakåt-lutningen bra, eller bör den ha olika accelerationslägen som när man snurrar på ljudet?
  - ”Tyckte ni det låt verklighetsstroget när ljudet flyttades runt (på en skala 1-10)?”
  - ”Tyckte ni det låt som ljuden befann sig mer utanför huvudet eller mer innanför huvudet?”
  - ”Tyckte ni det var lätt eller svårt att ta sig till ljudet? Varför var det lätt/svårt?”
  - ”Är det någonting annat ni tycker skulle kunna förbättras med träningsläget?”

**Testa spelet**

Testkör så det fungerar

Jag lyssnar samtidigt som de spelar, låt dem spela spelet 3ggr

Ev. **Anteckna det som verkar vara lätt/svårt**

**ANTECKNA:** Det längsta de kom på 3 försök

**P1:**
P2:
Frågor om punkterna jag har arbetat med, samt generella synpunkter (efter att båda har spelat):

"Vad tyckte ni om spelet rent generellt?"
- Hur underhållande upplevde ni att det var på en skala från 1-10?
- Tror ni att det skulle bli mer underhållande om ni fick spela fler gånger/öva mer?
- När ni spelade, kändes det som att ni verkliga ”var där i spelet” (skala 1-10)?

"Upplevde ni att någonting var dåligt/negativt eller förvirrande med spelet?"
- Märkte ni någonting udda (ex. konfirmationsljud när man inte klarat banan osv.)?

"Vad tyckte ni om spelets svårighetsgrad?"
- Vad tyckte ni om ordningen på banorna?
- Var spelet för enkelt eller för svårt?
- Vad tyckte ni om tiden per bana?

"Vad tyckte ni om banorna i sig?"
- Vad tyckte ni om ljudkvaliteten?
- Ev. Vilka upplevde ni hade sämst ljudkvalitet?
- Upplåste ni att ni kände igen ljuden i vissa banor mer än ljuden i andra banor?
- Ev. Vilka banor kändes mer verkliga?

"Vad tyckte ni om de andra ljuden i spelet (konfirmationsljudet, nedräkningar, game over med mera)?"
- Vad var bra/dåligt med dem?
- Skulle de kunna förbättras på något sätt?

"Övriga synpunkter på spelet?"
- Korta ”fångradien” där man får en highscore för att stegra svårighetsgraden? Dålig idé?

Testa banor de inte lyckades ta sig till (i synnerhet syntetiska banor där ljudets karakteristik förändras beroende på radien)

Börja på t.ex. level 11 (glöm inte att banga ny position innan)
Jag lyssnar samtidigt som de spelar (låt dem testa lite, max 2ggr)
Frågor om de syntetiska banorna (efter att båda har testat):

"Vad tyckte ni generellt om de syntetiska banorna jämfört med ljudfilsbanorna?"
"Vad är eurer åsikter angående att vissa av de syntetiska banorna förändras beroende på avståndet på andra sätt än att ljudstyrkan förändras?"
- Upplåste ni dessa banor som lättare eller svårare att klara?
- Upplevde ni förvirring av något slag?
- Tycker ni att de syntetiska banorna ska vara separata, eller kan de blandas med ljudfilsbanorna?

"Övriga synpunkter på de syntetiska banorna?"

TREDJE DELEN: IDEER FÖR ANVÄNDBARHETSTESTNING AV SPELET

Fick lite idéer av en grej jag programmerade:
Tiderna & alla highscores för varje bana sparas i textfiler
Ta reda på i vilken ordning banorna bör placeras (alla banor är skapta, men deras ordning är inte helt självklar i nuläget).
- Kan vara så att vissa banor som är lättare är placerade senare i spelet och vice versa
- Man skulle kunna låta flera personer få öva i 5-10 minuter och sedan spela spelet tio gånger utan tidsbegränsning, för att sedan räkna ut medelvärdet på tiderna och lägga banorna i rätt ordning
d. Efter detta kan man låta några blinda personer få öva i 5-10 minuter och spela spelet *med* tidsbegränsning och banor som är placerade i rätt ordning, för att få de slutgiltiga subjektiva åsikterna om spelet.

_Vad tycker ni om detta?_

i. Jag behöver alltså _två grupper_ nya blinda personer (minst två personer för varje test), så snart som möjligt (gärna nästa vecka).

**Tack!**

Hur snart kan vi ordna vidare användbarhets tester?

- 2 blinda för att spela alla banor utan tidsbegränsning
  - Gärna denna vecka eller nästa
- 2 blinda för att spela spelet med tidsbegränsning
  - Gärna nästa vecka eller veckan efter

Tack för all hjälp! Ni har varit suveräna.

---

**Användbarhets tester**

**Observationer**

_Ta fram anteckningspenna/papper_


Efter att du har spelat spelet kommer du sedan få _besvara några frågor_. Har du några frågor?"

**Perceptionstest**

- Öppna PRACTICE.pd
- Kontrollera nätverksverktyg och att IP-adressen stämmer
- Snurra runt ljudet och _kolla hur många rätt de får_ (5 ggr)

_Anteckna:_ Antal rätt

Förklara hur navigeringen fungerar och hur de gör för att få en score

Låt dem navigera fritt så länge de vill, men **högst 10 minuter**

**Låta dem spela**

- Öppna SOUND_HUNTER.pd
- _Döp om message-boxen för tiderna:_ level_times_name
- Säg åt dem att de nu ska spela spelet och klicka sedan på start

Medan de spelar

- Hur lutar de telefonen/navigerar?

_Anteckna:_ handrörelser (t.ex. lutar de telefonen för mycket, vilket håll?)
• Hur är deras huvudhållning?
   **Anteckna:** huvudhållning (försöker de luta/vrida på huvudet? Om de gör det, slutar de?)

• Är det någon bana som är helt omöjlig att klara?
   **Anteckna:** Vilken bana?

**Anteckna:** Hur många tips var jag tvungen att ge innan de klarade banan?

• *Efter varje omgång* är klar: **kopiera total highscore** till txt-filen **total_highscore_name**
   *Innan nästa spelare: Nollställ allting* (cleara tider/HS, sätt THS till 0)

**Intervju efter spelet**

**Bakgrundsfrågor**

Namn_______________

Hur gammal är du?
Svar_________________

Har du medfödd blindhet eller progressiv blindhet?
Medfödd – progressiv – ej blind

Har du på något sätt nedsatt hörsel?
Nej – ja, nämligen _________________

Har du spelat något ljudspel någon gång?
Ja - nej

Hur ofta spelar du andra datorspel/tv-spel?
Aldrig - väldigt sällan - sällan - då och då - ofta - väldigt ofta - hela tiden

Vilka ljuduppsättningar har du provat för datorspel?
Stereo-hörlurar – Stereo 2.0 – Stereo 2.1 – Surround(5.1/7.1) – Binaural HRTF – Övrigt

Vilka ljuduppsättningar har du provat för annat (exempelvis film)?
Stereo-hörlurar – Stereo 2.0 – Stereo 2.1 – Surround(5.1/7.1) – Binaural HRTF – Övrigt

**Spatialt ljud**

Känner du till begreppet Binauralt ljud eller HRTF?
Ja - Nej

Har du någonsin upplevt ”3D-ljud” genom hörlurar (exempelvis youtube-klippet ”Virtual Barbershop”)?
Ja – Nej

Beskriv kort vad du tänker när du hör uttrycket 3D-ljud!

---

**Lokalisering av ljud i spelet**

Ljudet befann sig på ett *annat* ställe än jag trodde
Aldrig - väldigt sällan - sällan – varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Ljudet befann sig på *samma* ställe som jag trodde
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Ljudet rörde sig *som jag ville* att det skulle när jag lutade iPhonen
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Ljudet rörde sig *inte som jag ville* att det skulle när jag lutade iPhonen
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Jag blev *förvirrad* när jag spelade spelet
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Att lokaliserare ljud i *sidled* var

Att lokaliserare ljud *framför/bakom* mig var

Ljuden lät som att de befann sig *ovanför* mig
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Ljuden lät som att de befann sig *under* mig
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Ljuden lät som att de befann sig *utanför* huvudet
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Ljuden lät som att de befann sig *inuti* huvudet
Aldrig - väldigt sällan - sällan - varken sällan eller ofta - ofta - väldigt ofta - hela tiden

Sättet ljuden rörde sig på i spelet lät *verklighetstroget*

Vissa ljud var svårare att lokalisera än andra (ringa in ett tal, 1= ingen skillnad mellan ljuden, 7=väldigt stor skillnad mellan ljuden)

(ingen skillnad)  1  2  3  4  5  6  7
(stor skillnad)

Om du upplevde en skillnad, varför tyckte du att vissa ljud var svårare att lokalisera än andra?

---

**Ljudens karaktär**

Jag upplevde att **kvaliteten** på ljuden i spelet var


Jag upplevde **brus eller andra konstiga ljud**


**Stördes du av brus eller andra konstiga ljud, det i spelet (om du upplevde det)?**


---

**Prestation**

Det kändes som att jag blev bättre på att **kontrollera** ljudet ju mer jag spelade

Inte alls – väldigt lite – lite – varken bättre eller sämre – något bättre – bättre – mycket bättre

Det kändes som att jag blev bättre på spelet för att jag **kom ihåg banorna**?

Inte alls – väldigt lite – lite – varken bättre eller sämre – något bättre – bättre – mycket bättre

Det kändes som att jag blev bättre på att **lokalisera** ljudet ju mer jag spelade

Inte alls – väldigt lite – lite – varken bättre eller sämre – något bättre – bättre – mycket bättre

Det kändes som att jag **fokuserade** mer på min hörsel när jag spelade


Det kändes som att jag **tränade** min hörsel när jag spelade


---

**Immiscion**

Jag tänkte på **omgivningen utanför spelet** när jag spelade


Jag **slutade tänka på tid och rum** utanför spelet när jag spelade


Det kändes som att jag **var där i spelets värld** när jag spelade

**Spelreaktioner**

Jag kände mig *lugn* när jag spelade

Jag kände mig *alert* när jag spelade

Jag kände mig *säker* när jag spelade

Jag kände mig *stressad* när jag spelade

Jag kände mig *glad* när jag spelade

Jag kände mig *ängest* när jag spelade

Jag tyckte spelet var *underhållande*

Jag tyckte spelet var *tråkigt*

Jag tyckte spelet var för *enkelt*

Jag tyckte spelet var för *svårt*

Vad var det som gjorde spelet enkelt eller svårt?

Enkelt, för att________________________________________

Svårt, för att________________________________________

Vad var det bästa med spelet?

Svar_________________________________________________

Vad var det sämsta med spelet?

_________________________________________________
Vad tror du skulle kunna förbättra spelet?

Jag skulle rekommendera spelet för andra
Ja – Nej