Examining visual cues for an auditory selective attention paradigm in virtual reality

————

Untersuchung visueller Cues für ein Paradigma zur auditiven selektiven Aufmerksamkeit in der virtuellen Realität

Bachelor Thesis

Pol de Lapparent

Immatriculation Number: 453432

April 29, 2022

Supervisor: Carolin Reimers
Supervisor: Alexander Heldring

Chair and Institute for Hearing Technology and Acoustics
Prof. Dr.-Ing. Janina Fels
RWTH Aachen University

Abstract

Over the last few years, virtual reality (VR) has emerged as a new technology that is allowing scientific research to investigate topics from another perspective. This is due to the possibility of creating more realistic scenarios in which all the variables are controlled. Thus, the conclusions taken from such research are more consistent since the environment can be much more controlled than without using VR. This thesis researched and validated a new visual cue in an experiment on auditory selective attention (ASA) experiment in virtual reality. The aim was to see how including a visual cue that participants had to search for affected their ASA, which would be a more realistic situation in a class scenario. In an already validated VR environment, changes were made to fit the purpose of this project. Four avatars were situated in four different directions in a classroom (front, left, back and right). One of the four avatars would clap at every trial, which would indicate the direction from which the auditory stimulus would later come from, and the participant had to search for it. When they did, they had to press a button, and then auditory stimuli coming from the avatar's direction would play. Another stimulus would play simultaneously from another direction to act as a distractor. Participants had then to indicate a characteristic of the target stimulus. To see how the participants' strategy worked in this new environment, the time it took to find the cue and the head tracking of the headset were saved during the experiment. The experiment was done with 10 participants. The experiment worked and proved that not all participants had the same strategy when searching for the cue. Participants would look at the cue for the front and sides positions when hearing the auditory stimulus. Whereas for the back cue, participants, when detecting that the avatar on the back was the one clapping, the participants would put themselves in a more comfortable position looking at the front to hear the auditory stimuli. Most participants' results were in a similar range, but there were different outcomes for two participants. One of them suffered cybersickness, and the other participant looked closer to the frontal position than any other participant.

Resum

En els darrers anys, la realitat virtual (RV) ha sorgit com una nova tecnologia que permet a la recerca científica indagar temes des d'una altra perspectiva. Això és degut a la possibilitat de crear escenaris més realistes on es controlen totes les variables. D'aquesta manera, les conclusions que s'obtenen d'aquestes investigacions són més consistents, ja que l'entorn pot estar molt més controlat que si no es fa servir la RV. En aquesta tesi es va investigar i es va validar una pista visual nova en un experiment d'atenció selectiva auditiva (ASA) en realitat virtual. L'objectiu era veure com la inclusió d'una pista visual que els participants havien de buscar afectava la seva ASA, es volia obtenir una situació més realista en un escenari de classe. En un entorn de RV ja validat, s'han introduït canvis per adaptar-lo al propòsit d'aquest projecte. Es van situar quatre avatars en quatre direccions diferents d'una aula (davant, esquerra, darrera i dreta). Un dels quatre avatars aplaudia a cada assaig, cosa que indicava la direcció de la qual provenia posteriorment l'estímul auditiu, i el participant havia de buscar-lo. Quan ho feien, havien de prémer un botó, i aleshores sonaven els estímuls auditius procedents de la direcció de l'avatar. Simultàniament, es reproduïa un altre estímul procedent d'una altra adreça perquè servís de distractor. Tot seguit, els participants havien d'indicar una característica de l'estímul objectiu. Per veure com funcionava l'estratègia dels participants en aquest nou entorn, durant l'experiment es va guardar el temps que trigaven a trobar la pista i el seguiment del casc de RV. L'experiment es va fer amb 10 participants. L'experiment va funcionar i va demostrar que tots els participants no tenien la mateixa estratègia a l'hora de buscar la pista. Els participants buscaven la pista per a les posicions davanteres i laterals quan escoltaven l'estímul auditiu. Mentre que per a la pista de darrera, els participants, en detectar que l'avatar de darrera era el que aplaudia, es posaven en una posició més còmoda mirant endavant per escoltar els estímuls auditius. Els resultats de la majoria dels participants estaven en un rang similar, però va haver-hi resultats diferents per a dos participants. Un va patir cibermalaltia, i l'altre participant va mirar més a prop de la posició frontal que qualsevol altre participant.

Resumen

En los últimos años, la realidad virtual (RV) ha surgido como una nueva tecnología que está permitiendo a la investigación científica indagar temas desde otra perspectiva. Esto se debe a la posibilidad de crear escenarios más realistas en los que se controlan todas las variables. De este modo, las conclusiones que se obtienen de estas investigaciones son más consistentes, ya que el entorno puede estar mucho más controlado que si no se utiliza la RV. En esta tesis se investigó y se validó una nueva pista visual en un experimento de atención selectiva auditiva (ASA) en realidad virtual. El objetivo era ver cómo la inclusión de una pista visual que los participantes tenían que buscar afectaba a su ASA, se quería obtener una situación más realista en un escenario de clase. En un entorno de RV ya validado, se introdujeron cambios para adaptarlo al propósito de este proyecto. Se situaron cuatro avatares en cuatro direcciones diferentes de un aula (delante, izquierda, detrás y derecha). Uno de los cuatro avatares aplaudía en cada ensayo, lo que indicaba la dirección de la que provendría posteriormente el estímulo auditivo, y el participante tenía que buscarlo. Cuando lo hacían, tenían que pulsar un botón, y entonces sonaban los estímulos auditivos procedentes de la dirección del avatar. Simultáneamente, se reproducía otro estímulo procedente de otra dirección para que sirviera de distractor. A continuación, los participantes tenían que indicar una característica del estímulo objetivo. Para ver cómo funcionaba la estrategia de los participantes en este nuevo entorno, se guardó durante el experimento el tiempo que tardaban en encontrar la pista y el seguimiento del casco de RV. El experimento se realizó con 10 participantes. El experimento funcionó y demostró que no todos los participantes tenían la misma estrategia a la hora de buscar la pista. Los participantes buscaban la pista para las posiciones delantera y lateral cuando escuchaban el estímulo auditivo. Mientras que para la pista de detrás, los participantes, al detectar que el avatar de detrás era el que aplaudía, se ponían en una posición más cómoda mirando al frente para escuchar los estímulos auditivos. Los resultados de la mayoría de los participantes estaban en un rango similar, pero hubo resultados diferentes para dos participantes. Uno de ellos sufrió ciberenfermedad, y el otro participante miró más cerca de la posición frontal que cualquier otro participante.

Contents

1 Introduction

The human ability to concentrate on a specific sound source is auditory selective attention (ASA). It is a topic that has been researched for many years, and in recent ones, it has been studied in a Virtual Reality (VR) environment. Working in VR allows to have a better control of the variables happening in the experiment.

The topic of this thesis takes the continuation of previous work done at the Institute for Hearing Technology and Acoustics at the University RWTH in Aachen. The last paper of all this research focused on using two-dimensional exogenous visual cues set in a fixed place of the virtual world to direct the participant's attention to a given direction to hear a sound source. [13] This setting was used to show that a previous validated child-appropriate paradigm on auditory selective attention could be transferred into VR. It succeeded but also concluded that it wasn't 100 per cent realistic since no head movement was involved.

Therefore, this thesis goes further into the topic, taking Head Movement from the start. The paradigm of this experiment uses movement from avatars to guide the participant's attention towards a specific position. Then an auditory stimulus is played from that direction, and the participant needs to detect it. Hence, the visual cue in this experiment is no longer in a fixed place of the virtual world but four different possible places (Front, Left, Right, Back). The visual cue in the previous experiment was a two dimensional Lego cue indicating the direction to focus. Whereas, in this experiment, there are four avatars created with Mixamo [1], where the visual cue is the avatar that is doing a unique animation (in this case, clapping).

The thesis is structured in the following way: First, in chapter 2, the state of the art presents each of the critical concepts of this thesis that will then be used further in work. Then, there is chapter 3 with implementation, which explain how the listening experiment was planned and conducted and each of the elements that made it possible. Finally, chapters 4 and 5 give the results, a discussion on them, and the conclusions and further potential research that can be done in the future.

2 State of the Art

This chapter gives an introduction to each one of the main concepts of this thesis. This is important because it provides an initial overview of ideas that will be then mentioned in both chapters 3 and 4.

The first section will cover binaural hearing; the second one will be auditory selective attention. While in the third section, virtual reality concepts will be discussed.

2.1 Binaural Hearing

The capacity to hear with both ears by processing information through pressure waves is known as Binaural Hearing. This critical ability that humans possess allows them to pinpoint any incoming sound; this is very useful since people's vision is limited to a scope but not hearing capacity. This chapter will present an overview of the different characteristics that encompass binaural hearing. First, the concept of Head-related coordinate system will be introduced, then tools used to detect sound will be discussed. Finally, the head-related transfer function (HRTF) and the Headphone Transfer Function (HpTF) will be explained.

2.1.1 Head-related coordinate System

On a psychoacoustic experiment, the listener perceives the sound from a first person's perspective. Therefore, the 3D terminology of the sound's origin needs to be adapted to this case. There are three planes in the Head-related coordinate system: the Horizontal plane, the Median plane, and the Frontal Plane. Those three planes are defined by an axis that divides the head into two symmetric parts and the interaural axis, which spans from one ear to another. In the following picture (figure 2.1), the planes and the axis that form the Head-related coordinate system can be seen.
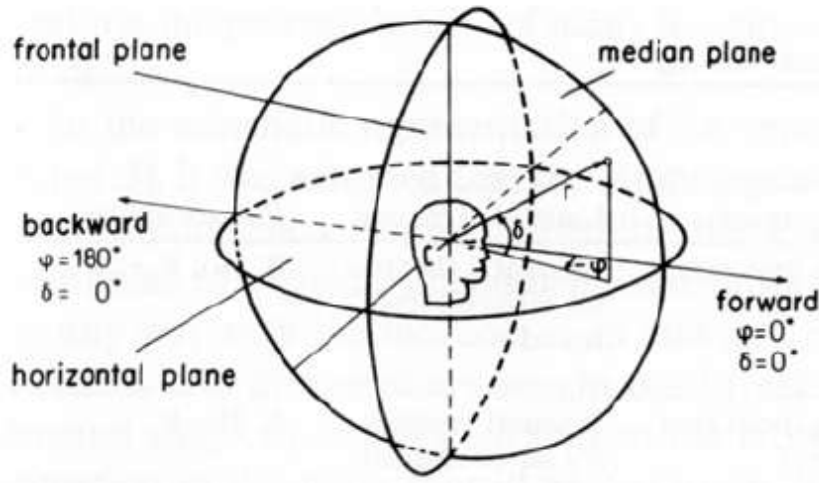
Figure 2.1: Head-related coordinate System [7]

2.1.2 Interaural time difference (ITD) and Interaural level difference (ILD)

Humans use both Interaural time difference (ITD) and Interaural level difference (ILD) when wanting to detect the position of a sound source. ITD uses the time difference between the arrival to each ear for a pressure wave.

Whereas for ILD, it is used how the wave pressure enters the ear (see figure 2.2). This stems from the idea that different reflections and diffractions will play out depending on how the wave enters the ear.
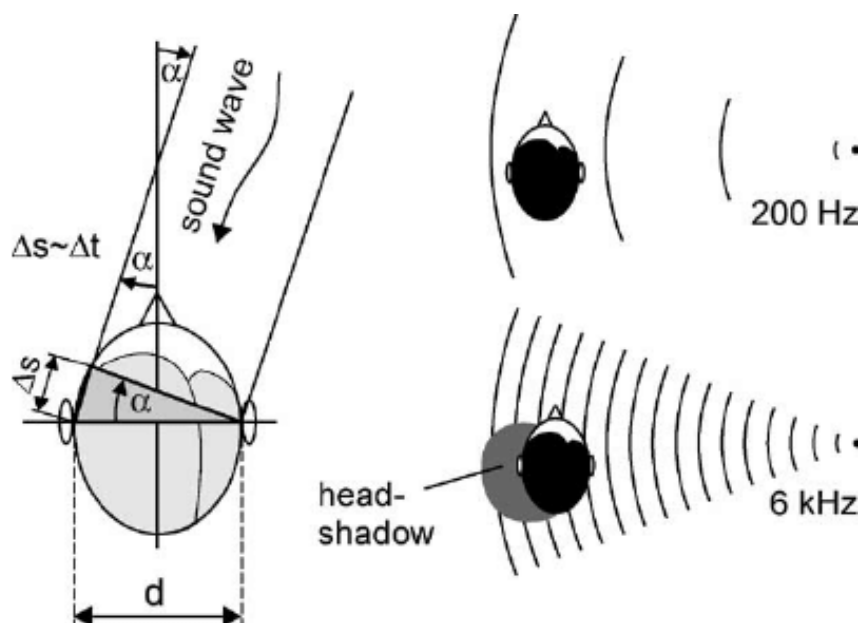

Figure 2.2: Interaural level difference [18]

2.1.3 Duplex Theory

Lord Rayleigh introduced the Duplex theory in 1907, and it said that both ITD and ILD work predominantly in specific frequencies. As seen in the explanation of ILD, it is frequency-dependent, and its use in low frequencies is minimal, whereas, in higher ones, it is genuinely effective. Anyway, since most sound waves have both high and low-frequency components, both ITD and ILD will overlap and be used to detect the whereabouts of sound sources. The liability of this theory is that it does not take into account points in the median plane, which is equidistant from both ears and where ITD and ILD don't play a role.

2.1.4 Head Related Transfer Function (HRTF)

Since the use of the Duplex theory is incomplete with sound sources in the median plane, there is a need to have a tool to understand how do people perceive each sound wave. With this idea comes the head-related transfer function (HRTF), which defines each ear's filtering to the sound to perceive the information. It describes the spectral characteristics of the sound when it is measured in the tympanic membrane. [2] The HRTF is unique to every human since it takes into account the different amplifications and attenuations that the pressure wave has in the ear canal. This function is unique since each human posses difference in their ear, which makes the HRTF in one way or another.

The HRTF is acquired by taking the Fourier transform of the head-related impulse response (HRIR), which is the filtering applied to the sound wave but in the time domain. It can be done individually but it is very costly, so in this thesis an artificial head was used. Of course, by using an artificial head and not a individualised HRTF for each participant there some problems that appear for the participant when localizing a sound source. The main one would be confusion between front and back sounds.

2.1.5 Headphone Transfer Function (HpTF)

To achieve a perfect binaural reproduction in the experiment, there has to be a certainty that the binaural signals that the headphone plays is consistent. The issue is that different headphones have different transfer functions depending on how they are designed. So a Headphone Transfer Function is (HpTF) measured to equalize the headphones used in the experiment. The method used to get the headphone equalization was introduced by Masiero and Fels [11].

The procedure to calculate the HpTF was the following. Participants had to put small KE3 microphones into their ears, followed by the headphones later used in the experiment. Then a Matlab code was played that measured six times each ear using sweep signals. In between the six measurements, the participant needed to take out and put the headphones again. The head movements were not considered since this measurement was done in static.

2.2 Auditory Selective Attention

Auditory selective attention (ASA) is a key skill that all of us possess which entails focusing your attention into a particular sound source. Thought since it is an ability that has daily usage, it tends to be used without even realizing sometimes. This concept has been research for many years since it can be applicable in many situations. It was first describe in the research done by Cherry referring it as the cocktail party phenomenon [4]. In this research, participants had to focus on a stimulus coming to one of the two ears discarding another stimulus from the other ear. The conclusion of the project was that people were able to not process the full information on the ear they were not suppose to focus but still manage to figure out abrupt changes in the distractor stimuli. Overall ASA was described as the ability to focus on a particular sound source given specific audio characteristics. First it was shown, that semantically relevant information such as one name is not suppressed. This was done with diotich and dochotic hearing, where no spatial information on the sound was provided. One of the first ones to investigate actual binaural hearing were Fels, Koch et al [8]. Then came the work of Oberem et al [12] which took a new approach by extending the paradigm in a more natural and realistic environment This is also why the work of Oberem et al is so important. It demonstrated that when the stimuli is presented binaurally and not dichotically, both error rates and reaction time went down. The next work that has a significant importance in this introduction of ASA is the work done by Loh, Fels et al [9] which based their research in validating a new paradigm in which the comparison of intentional switching in ASA was studied. Further studies from Reimers, Loh, Fels et al [13] took this already 'previously validated paradigm and took in into VR. This was done by creating a new VR environment of a typical classroom and an experiment like the one with Loh, Fels et al was carried but in VR. Results from the experiment were very similar from the already work done but not in VR, hinting that analysing ASA in VR is an appropriate way to go. This experiment continues with this path of research by introducing a new visual cue into the experiment designed in Reimers, loh, Fels et al.

2.2.1 Visual Cues

So not only sound waves can impact ASA, but also visual phenomena. In the experiment done in this thesis, visual cues are used to direct the participants' attention to a particular direction. But what is understood by visual cues? In visual perception, the eye receives sensory cues, which are visual cues that are in the form of light that is processed by the visual system. So it is referred to any visual phenomenon that is then later processed as information by the brain. Humans tend to focus on what they sense, and sight is 80 per cent of all the information perceived. So a study on how visual cues affect Auditory Selective Attention does make sense since there is a significant dependency on this sense.

2.3 Virtual Reality

With each advancement in technology, new tools to go further into other research fields emerge. And this is especially the case of Virtual Reality (VR), which helps recreate more controlled environments in which each characteristic and variable to shape the experiment in the way we

want can be controlled. This allows that the conclusions done by such research give a new dimension since the imprecision is reduce to its minimum.

VR refers to a computer-generated environment with scenes and objects that appear genuine, thus giving the user the impression that they are immersed in their environment [6]. An individual can experience this environment by wearing headgear or a helmet or in a CAVE system. Through virtual reality, immersion in video games is now possible as if being the first character, learning how to perform heart surgery, and improving sports training to maximise performance. In this thesis, Virtual Reality allows us to recreate a controlled class environment in which ASA can be experimented.

2.3.1 Head Tracking

The position and rotation of the HMD headset are saved each time the visual cue starts playing. When the participant moves in the VR environment, this tracking is what allows to see where is the participant, and what is their rotation. From this it can be seen also their rotation, and therefore, where are they looking at. The Virtual Reality setting records both the position and rotation of the headset. Since the participant will be sited, there is only the need to analyse the rotation of the headset, which in this case is the rotation. Unity saves the rotations, not as Euler angles but as Quaternions, since in game design, its benefits surpass its limitations. So before any analysis has been done, a conversion to Euler angles is needed to visualise the data better.

2.3.2 Virtual Agents

For this project, four avatars are used as the visual cues of the experiment. In this subsection, different possibilities for making the avatars will be discussed.

2.3.2.1 Blender

The first and most obvious possibility was to create an avatar from scratch with Blender [3]. Blender is an open-source software that allows the creation of many three dimensional graphics. Working with Blender means to work from scratch and many time as well as skill is required in order to develop an avatar to the specifics wanted to fit a psychoacoustics experiment accordingly.

Nevertheless, it added too much complexity to the project. First, a lot of experience designing avatars from scratch was needed. But most important because after some research, getting a satisfactory result for the psychoacoustics experiment would have required a considerable amount of time.

2.3.2.2 MakeHuman

Other applications were reviewed that already had avatars created, and only slight modifications were needed. One of these was MakeHuman [16], an open source software used to create 3D avatars. Having at its base a basic human mesh, it allows to create avatars by modifying a couple of features each time. The problem came when it was added to the VR environment; it was a

clear mismatch appearance. What's more, MakeHuman only gave the avatar and the animation needed to be created, so it wasn't the way to go either. In figure 2.3, a potential Makehuman avatar is shown.
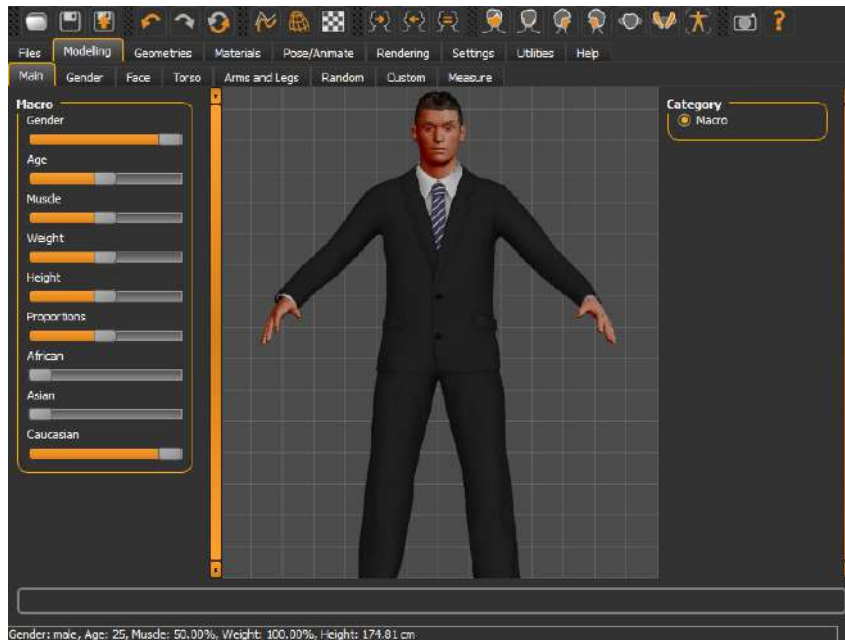

Figure 2.3: Makehuman avatar not choosen for the experiment [16]

2.3.2.3 Mixamo

As a third option, Mixamo was investigated [1]. Mixamo is an online avatar creation tool by Adobe which contains lots of characters with different animations ready to use. Both the avatars and animations looked quite serious, and when bringing them into the VR environment, they looked adequate with the experiment's objectives. There are many possible characters inside Mixamo, but the Y bot was selected since it looked neutral and would be the best choice for experiment purposes.
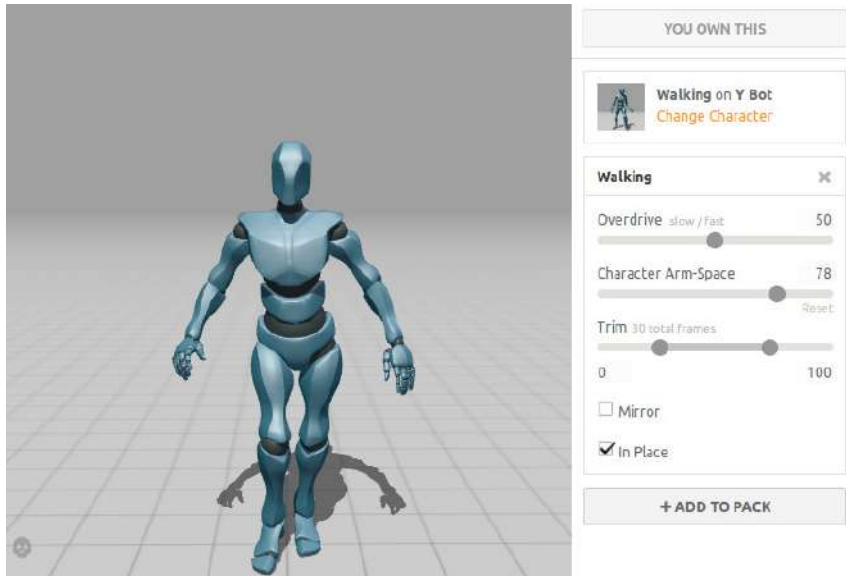
Figure 2.4: Y bot used in experiment [1]

3 Implementation

A listening experiment was done to see how people's attention shifted and how their strategy was doing while using a visual cue that moved. The following chapter will discuss the implementation elements that made the experiment possible.

3.1 Participants

As the auditory stimulus material consisted of spoken animal names in German language, the participants needed to have good German skills (B1) to distinguish effortlessly between the eight animals played in the listening experiment. All in all, 10 different people participated in the experiment (mean = 22.9 years, Standard Deviation = 0.876 years, 9/1 male/female). They all had normal hearing between 250 Hz and 10 kHz (within 25 dBHL) and normal vision. To check their normal hearing, audiometry was done with the software Ear 3.0, which uses a pure-tone in each ear, increasing the SPL when the listener detects it. None of the participants had done a listening experiment before, and all of them gave their consent to being there and using their data for research purposes.

3.2 Room Equipment and Set-up

The listening experiment was done in a research lab in the Institute for Hearing Technology and Acoustics at RWTH Aachen University. Inside this research lab, a hearing booth of $2.3x2.3x1.98m^3$ is placed with VR material like headset and controller to do the experiment. This helps to have a closed area with better acoustic conditions, which help reduce at its minimum other variables that we can't take control of within the experiment.

Inside the hearing booth, there is a chair in the middle position, two SteamVR lighthouse Base stations, headphones, one headset HTC Vive Pro Eye with two HTC Vive Controllers. During the experiment, the participants are sitting on the chair and are free to move their head. The Virtual Reality Base stations detect and monitor both the Headset and the controllers. The participant, while sitting, wears both headphones and headset while they control each of the controllers with their hands. They are told in the instructions that the only button used during the experiment is the trigger situated in the back of the controller.

3.3 Stimulus Material

During the experiment, stimulus material were or by a 24 years old woman or a five years old boy. The duration of this given stimulus is no longer than 600 ms since that was the intended duration. It was manipulated to be 600 ms using a time stretching algorithm that doesn't influence the pitch. The possible Stimulus, as seen in table 3.1, are Katze (cat in English), Schlange (snake in English), Ratte (rat in English) and Robbe (seal in English) for the non-flying animals. Eule (owl in English), Ente (duck in English), Taube (pigeon in English) and Biene (bee in English) for the flying animals.

Stimulus Material	
Flying Animals	Eule, Ente, Taube and Biene
Non-Flying Animals	Katze, Schlange, Ratte and Robbe

Table 3.1: Stimulus Material

3.4 Binaural Sound Reproduction

As stated in the State of the art of this thesis, a Head-Related Transfer Function (HRTF) is needed to ensure that every participant hears, in the same way, all of the stimuli presented above. For each of the participants of the experiment, to get their HRTF, their head was measured, and their Headphone transfer equalisation (HpTF) was measured. The HpTF was calculated using the method introduced by Masiero and Fels [11] using the ITA-Toolbox in matlab done by Dietrich et al [5].

In this project, the playing sound comes from different directions. The sound stimulus can come from four different possible positions for this experiment. Being those, front, left, right and back as seen in Figure 3.1. It is important to note that the direction of the sound is dynamic.
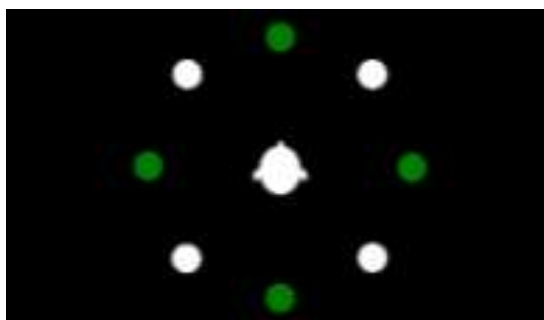
Figure 3.1: Possible directions of the sound sources [10]

3.5 Cognitive task

The core part of the experiment, in which the participant is asked to play in a given scenario, is called the cognitive task. The participant, under some circumstances, has to differentiate from

two sound sources (Target Stimuli Distractor Stimuli) and pinpoint if the Target Stimuli is a flying or non-flying animal.

3.6 Experimental Procedure

To design this experiment, a new visual cue was designed and implemented into the experiment. In this section the experimental procedure will be explained.

3.6.1 New visual cue

So for this experiment, the goal was to introduce a new visual cue to the already previously validated paradigm. After much research, it was decided to have four avatars in the four possible target position of the stimulus, and that they somehow indicate the participant that the stimulus would come from there. This was to go further into the research and get into a more realistic paradigm. Since normally in a class environment, people tend to be distracted by other directions than the frontal one. After the analysis of the virtual agents in subsection 2.3.2, Mixamo was the software chosen for this experiment. The main reason behind was that it merged the best with the already created VR environment. Furthermore, it also brought animations so it made everything easier. The animation chosen was the clapping one, since it is a common movement used in a class environment. In next figure 3.2, the avatar that was in the front position is clapping.

Figure 3.2: Front avatar clapping

Overall, the investigation consists of a total of 288 trials in 6 blocks (336 trials in 9 blocks taking into account the also the trial blocks).

3.6.2 Trial Structure

In this experiment, the participant is sited in a Virtual Reality classroom in which four avatars are sitting in the positions where the sound is playing from. The participant at each trial has to look around the classroom to see which of the four avatars is clapping. Once this is done, they press one of the Trigger buttons in one of the controllers, and then two sounds will be played simultaneously. The first one, the Target Stimulus, will come from the avatar's position that is clapping. In contrast, the second one, the Distractor Stimulus, will come from another random place to try to confuse the participant. Then it is required for them to press the trigger which has the picture of the flying animal or non-flying animal depending what they heard from the target position.

After the input of the answer, a feedback screen appears on the blackboard of the virtual reality classroom. Two possible images can appear whether the answer was the correct one or not. If it were, a lego smiling face would show itself saying "RICHTIG" (right in English), and if it were incorrect, a lego sad face would appear saying "FALSCH" (false in English). At the end of each of the six blocks, a resume of all the answers of the block will appear. Depending on their error rate, a rating will be given in stars (between 3 to 5). It follows the following pattern:

- Error rate $< 0.2 = 5$ stars

- Error rate $< 0.4 = 4$ stars

- Error rate $> 0.4 = 3$ stars

After the feedback of the block is given, a message will show, indicating the participant to take a slight pause before continuing the experiment. This is much more important in this experiment since they have to be constantly could generate more easily into cybersickness.

The following picture shows the virtual reality classroom designed, in which the central chair pictures where the participant is sited during the whole experiment. Then, a table summarises the trial structure with every element and its known duration. Of course, for both the searching of the visual cue and the answering of the stimulus, there is no limit since those are the ones that will be then studied.
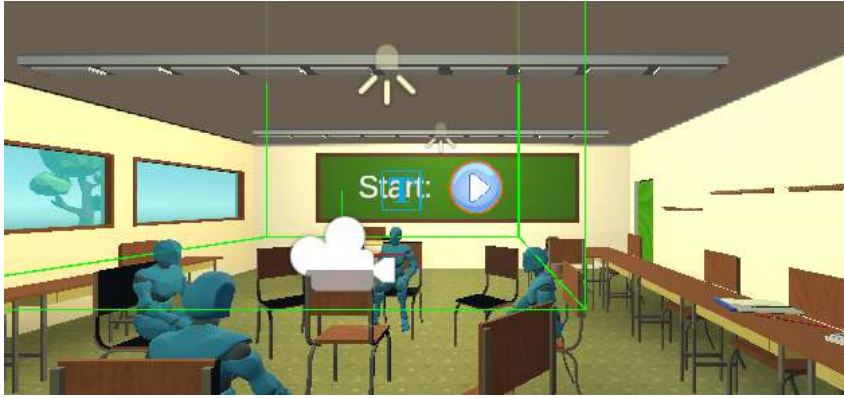
Figure 3.3: Virtual Reality Classroom

3.7 Experimental Design

To determine how people would react in this new environment, it was necessary to set and fix the different dependent and independent variables before the start of the experiment. The experiment has a total of three independent variables and four dependent variables. Of the independent ones, there were 16 combinations on factor levels called conditions. Each condition has been repeated a total of 18 times to have a clear picture of how the independent variables affect the dependent ones. An overview of the whole experimental design will be layered out in the following subsections.

3.7.1 Independent Variables

The three different independent variables are the following: Positioning of the Visual Cue, Transition and Congruence.

3.7.1.1 Positioning of the Visual Cue

This experiment brings a new dimension by telling the participant to rotate themselves searching for the visual cue, whereas in others, they didn't need to move since everything happened in the front position. So the Positioning of the cue is one independent variable with four levels: front, left, right and back.

3.7.1.2 Congruence

Congruence is the concept that relays how the target and distractor stimuli are alike. This independent variable is used to understand how people pay attention to irrelevant information. There are two levels in this variable: congruent and incongruent. Congruent will be in cases in which both the target and distractor stimuli belong to the same category (e.g. the target is Ratte and the distractor is Katze) where both animals fit in the flying category or the non-flying one. In contrast, incongruent is when the distractor stimulus is different from the target one. Or the target is a flying animal, and therefore the distractor is non-flying or the other way around.

3.7.1.3 Transition

The final independent variable used in this experiment is transition. This concept is used to see how the block will be played out. This variable has two different levels, switch and repetition. Both variables are self-explanatory since they relate to the idea to repeat or switch from where the target stimuli were coming from. So if the previous target stimuli were on the left, and now the variable transition is switched, the subsequent target stimuli will be in another position. Whereas in repetition trials, the target stimulus comes from the same direction in the next trial.

3.7.2 Dependent Variables

The dependent variables in this experiment are the following: Error rate, Reaction time, Time to find the visual cue, and tracking the head movement.

3.7.2.1 Error rate

The error rate (ER) is the variable that determines how many errors are the participant doing during each of the six blocks in the experiment. For matters of time, there has not been an evaluation regarding the results of the ER, but this could be done in the future. But it was still measured to give credibility to the data acquired from other dependent variables. Since the experiment presents a new dimension, it was necessary to assess that participants still had good results. If not, the conclusion taken from other variable analyses wouldn't have much value.

3.7.2.2 Reaction time

The time it takes from hearing the stimuli to indicate what type of animal was; is what the variable reaction time records. In the experiment, the variable was measured for good practice, but wasn't evaluated since this is not of this thesis interest.

3.7.2.3 Time to find the Visual Cue

From the time that each trial starts until the participant presses one of the two triggers buttons, some time passes, depending on where the target direction is.

3.7.2.4 Tracking of the Head movement

To find which avatar of the four is moving, the participants need to turn their heads. Also, although the head tracking of the headset will be recorded during the whole duration of the experiment, only a subset of all this tracking will be later analysed. This subset of data on the time when the participant presses their trigger button and stimuli play out. This is to analyse the strategy people take to concentration. It is important to note two critical aspects to the Head movement data tracked in the experiment. Since participants do not move from the chair and only turn, only the rotation of the yaw has been taken into account (axis y). Like most

game engines, Unity VR uses quaternion as their mean to express rotation. Using a Matlab code, the data from the quaternion has been translated into Euler angles to ease the analysis of the further conclusions of those. Secondly, since the participants have the same strategies to look into both right and left sides, only the imaginary part of the quaternion was taken into account.

3.7.3 Hypothesis

This experiment was done how would people react in a more realistic classroom setting so the first idea that came while designing it was if participants would look at the target position while listening to the target stimulus. The hypothesis before the start of the experiment was that participants would indeed look at the direction of the target stimulus but from the back position. Since turning around in a classical chair is uncomfortable, the hypothesis was that people would detect the back cue and then return to a more comfortable position to hear the stimulus. In the results chapter, in the discussion of the head tracking data it will be discussed deeper this hypothesis and see if the results match it or not.

4 Preliminary Timeline

In this chapter I will comment both the Initial Timeline I had when starting the project and then which has been the final one. At the end of this chapter there is a Gantt chart where all the information above will be visualized.

I would say that what impacted the most between my first timeline and the final one is my lack of knowledge to know how much certain parts of the projects would take. My home university had two possible delivery dates, one in late January and another one open until end of May. Even if RWTH calendar starts one month later than my home university for the winter semester I had hopes I could make up the time and arrive to this final deadline. Unfortunately, I realize pretty quickly that being able to finish in the time frames I had set myself was unrealistic. The Literature Review and Understanding of the Procedure wasn't too much of a problem with my supervisor's help but when I got into Implementation it got increasingly difficult. When I decided what I wanted to do in the experiment, using avatars as a Visual Cue, I had to learn how to do what I was intending to. I had overconfidence on myself I would say, and I thought that modifying the previous VR environment to my needs wouldn't be too complicated given the knowledge I had. One of my biggest struggles in this part was understanding how Unity works, which I had never used before. I would certainly say I enjoy this part of the thesis in which I had to do a lot of self-education on the topic but it was quite frustrating the first weeks. I wanted to consistently try to correct the errors trough coding like I have done in the majority of my Bachelor's but it wasn't until I had a complete understanding on how a Game Object in Unity works I could get a result without any error. Thereafter, I was quite stressed because since I though that I had spent much more time than I had expected I worried other parts of the project would also go wrong very easily. Luckily when I finished my VR environment and the coding of the experiment, no major problems emerged. I could present my experiment to the Institute researchers that gave me feedback on how they though my experiment planning was. I was then able to do the experiment, record the data, do the evaluation and write the thesis without much of an interruption.

- Month 1:
    - Literature Review
    - Understanding the Procedure
- Month 2:
    - Implementation
    - Experiment Design
- Month 3:

– Experiment Coding

- Month 4:

    – Mid-term presentation

    – Done Coding

- Month 5:

    – Experiment

    – Data Analysis

- Month 6:

    – Data Analysis

    – Writing

    – Proof Reading

    – Finalizing the Thesis

In the following four pictures, there is represented the Gantt chart I have for my thesis. The blocks that are in green represent the duration that were supposed to last and in yellow is the amount of extra time that was needed to complete such task.
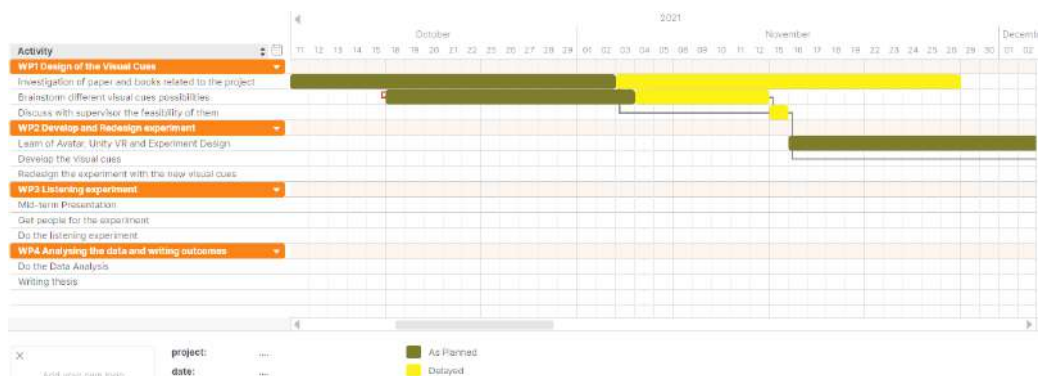
Figure 4.1: Gantt October-November
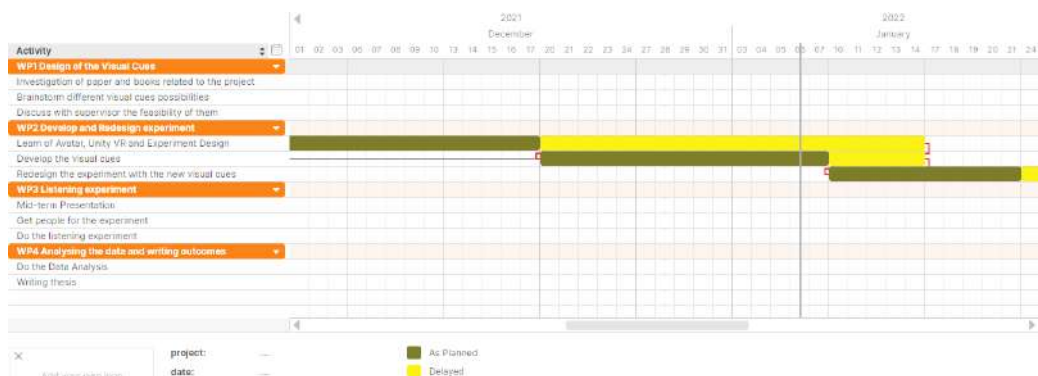
Figure 4.2: Gantt December-January
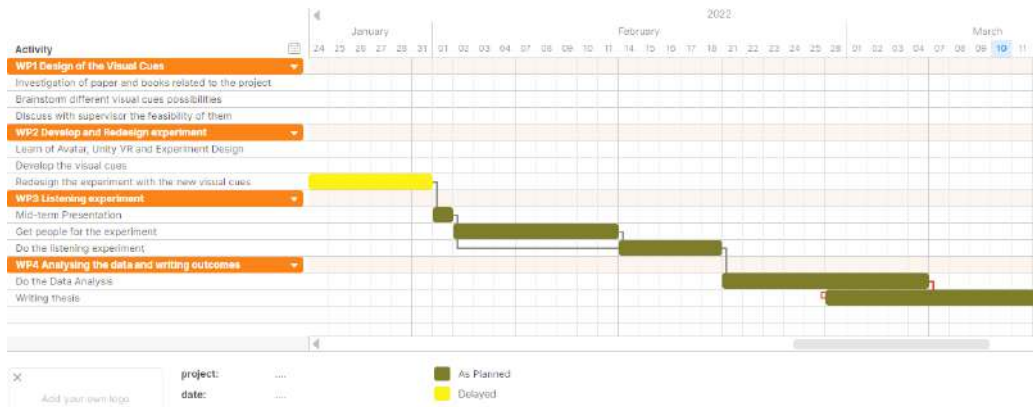
Figure 4.3: Gantt February-March
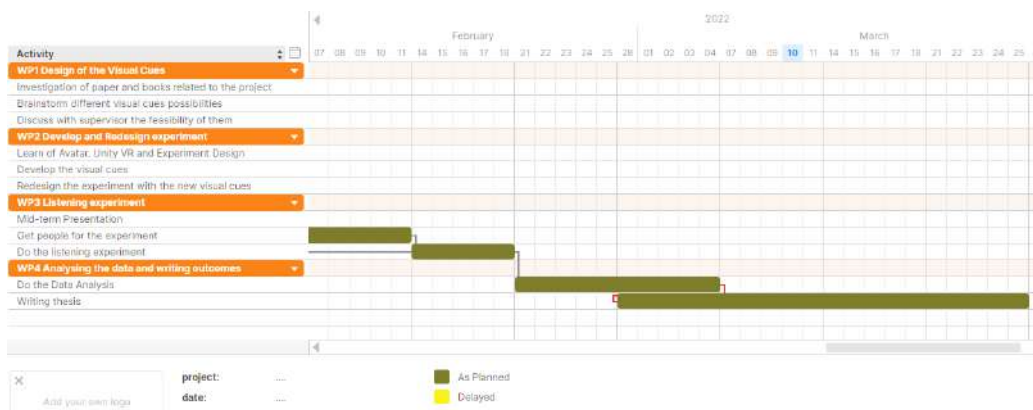
Figure 4.4: Gantt March

5 Results

In order to analyse the results from the VR experiment, linear regression was used to see how participant's strategies would shift during all the blocks. Two studies of linear regression were studied, in which the block number acted as a predictor for both the time of finding the visual cue and the rotation of the participant headset. But before all, the data needed to meet certain assumptions so that the linear regression would be valid. The software used for the analysis of the data is IBM SPSS.

5.1 Assumptions for Simple Linear Regression

Several assumptions need to be met to analyse the data with linear regression. In this section, the assumptions discussed will be outliers, linearity of the data and homoscedasticity. [14]

5.1.1 Outliers

To get some consistent data to work on with the analysis, it was necessary to make sure that no data was too inaccurate to affect the linear regression negatively. Looking at the values of the time in finding the visual cue, a variety of minimal values can be seen. This could happen for several reasons, but the most likely would be an accidentally double tap on the trigger button or a cybersickness during the experiment that led to not ending the experiment forcefully as early as possible. So, all values beneath 150ms [17] have been taken out of the data set. Then, other data can hinder the analysis if they are far enough from the mean. All of the data that was more than two standard deviations away from the mean was excluded for the following analysis. This particular set of outliers will only be taken out of the data set of the time finding the visual cue. This is done this way because, in the analysis of the head tracking data, it is necessary to evaluate the strategy that the participants took during the whole experiment event if there were cases in which it doesn't follow the linear model. The following Table 4.1 shows how many values were used for each position and block in the linear regression. The value of 120 was the number of values each block had at each position before taking out the outliers. All in all, 10.28 per cent of the data were outliers.

Blocks	Target Position			
	Front	Back	Left	Right
Block 1	109/120	107/120	105/120	107/120
Block 2	110/120	104/120	111/120	113/120
Block 3	106/120	109/120	104/120	110/120
Block 4	105/120	107/120	107/120	111/120
Block 5	109/120	104/120	111/120	110/120
Block 6	105/120	108/120	105/120	107/120

Table 5.1: Values used without outliers

5.1.2 Linearity of the data

To analyse linear regression, it is critical to see whether the data to be analysed has a linear dependence between the prediction and predictor value. The linearity of the data assumption was checked with a normal Probability-Probability plot (P-P) of Regression Standardized Residual. The theory states that for this assumption to be accurate, the data represented on the graph has to follow the diagonal line [14]. There exists room for some deviation of the data, but the pattern of it has to stay the same and tend to the values of such line. Further ahead, two P-P plots, figures 4.1 and 4.2, can be seen of two subsets of the Head tracking and Visual cue data set. The other P-P plots can be found in the Appendix (Figures A.1, A.2 and A.3 for the head tracking data and Figures A.4, A.5 and A.6 for the time in finding the target cue data).
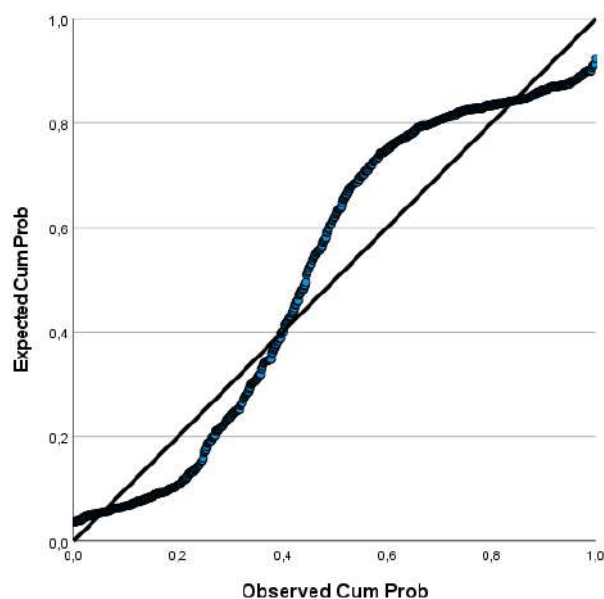
Figure 5.1: P-P plot of the head tracking data when the Target Position is the right one
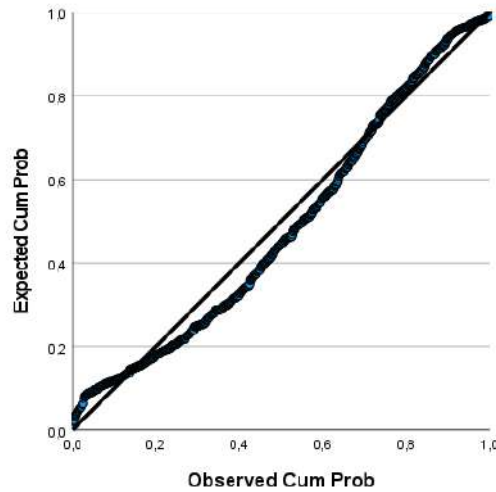
Figure 5.2: P-P plot of the time finding the visual cue data when the Target Position is the left one

5.1.3 Homoscedasticity

The assumption of homoscedasticity checks if all the errors between all the values in the model are similar or not. The error mentioned here refers to the one between the actual value and the predicted ones by the model. If a data set is homoscedastic, its scatter plot of its residuals will show a shut-gun result, with values in an unorganised pattern. As seen below, a scatter plot of the residuals of one of the analyses of the visual cues, and one of the head tracking analyses has been done. In each block, the data represents an unorganised pattern and, therefore, checks the homoscedasticity assumption in both cases[15]. The study of the homoscedasticity of the other subsets can be found in the Appendix (Figures A.7, A.8 and A.9 for the head tracking data and Figures A.10, A.11 and A.12 for the time in finding the target cue data).
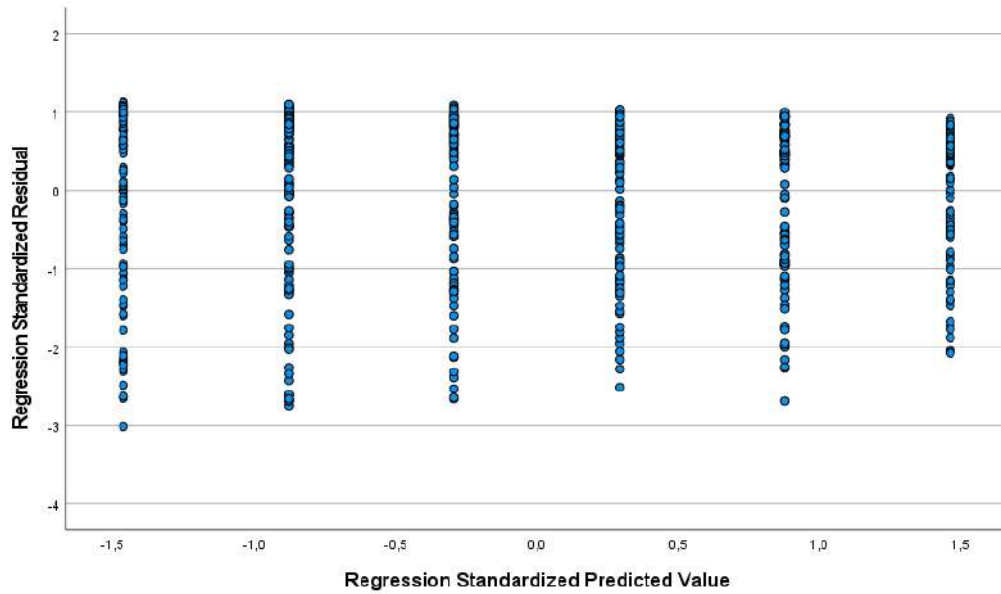
Figure 5.3: Scatter plot of residuals of the head tracking data when the Target Position is the back one
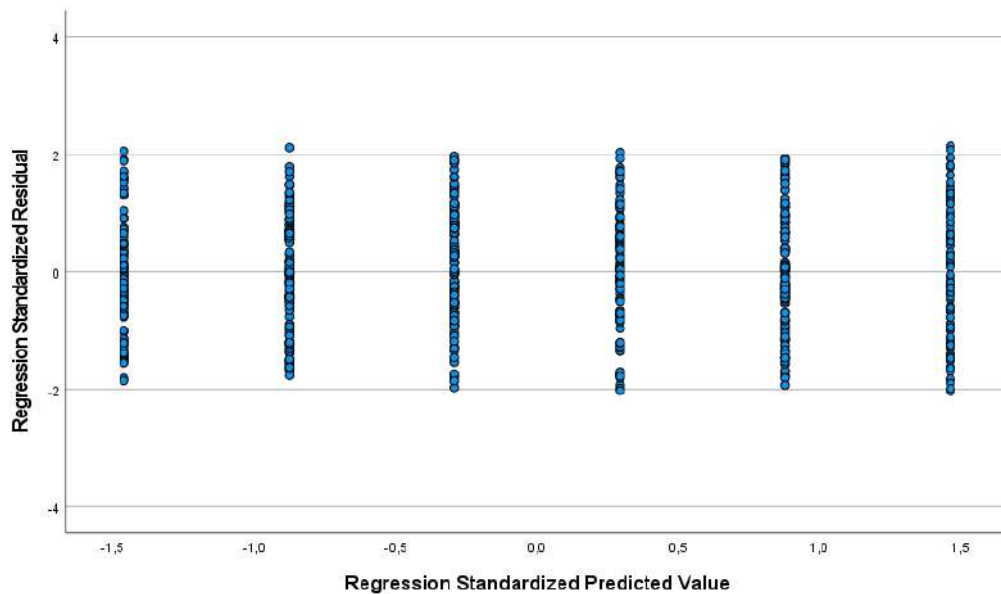
Figure 5.4: Scatter plot of residuals of the time finding the visual cue data when the Target Position is the back one

5.2 Analysing Time in finding Visual Cue

As stated before, one of the two dependent variables of this experiment is the time taken by the participant to see which of the four avatars was the one clapping and, therefore, where was the target stimuli going to come from. After the data was separated between each target position and checking the assumptions in subsections 4.1.1, 4.1.2 and 4.1.3, a linear regression model was done over the data. The following table 4.2 shows the different values of the linear regression

analysis. The time values are in milliseconds.

Linear Regression Data	Cue Position			
	Front	Back	Left	Right
Mean(ms)	1128.38	2261.28	1470.50	1528.51
Unstandardized B (Constant) (ms)	1316.821	2617.947	1873.072	1906.091
Unstandardized B (Target Position) (ms)	-54.129	-101.835	-114.944	-108.067
R	0.194	0.171	0.252	0.214
R^2	0.038	0.029	0.064	0.046
Adjusted R^2	0.036	0.028	0.062	0.045

Table 5.2: Results for the linear regression on the time to find the cue in relation to the cue position

As seen from the data in the table, there are some logic results have come through after the analysis and some that not so much. First, there is an obvious longer task when searching for the back avatar than it is to search either one of the sides and even more than the one at the front. This initial first belief is confirmed by the Unstandardized B (Constant), which shows the value that the regression line hits the Y axis. Indeed, it can be seen that the highest value among the four possible positions is the Back one, with an Unstandardized B (Constant) of 2617.947 ms. Also, both sides' avatars have a similar value in this area, which should be understandable since not much more movement is involved. And finally, there is the frontal cue which has a much lower value than any other side.

From the other side, it can be seen with the R^2 and Adjusted R^2 that the linear regression model doesn't appear to be able to predict all the variation of the data. In order to be that way, the values of this two components of the model should be closer to one. That would then indicate that the regression model calculated would fit the data. This information shows the linear regression model doesn't explain all the variation of the data that there is around its mean. The linear regression plots can be found in the Appendix (Figures A.13, A.14, A.15 and A.16). For this reason, the further analysis of the results, will be conducted with box-plots where the data can be seen in a more visual manner. In the following box plots (Figures 4.5, 4.6, 4.7 and 4.8) the results of each position in each block can be seen.
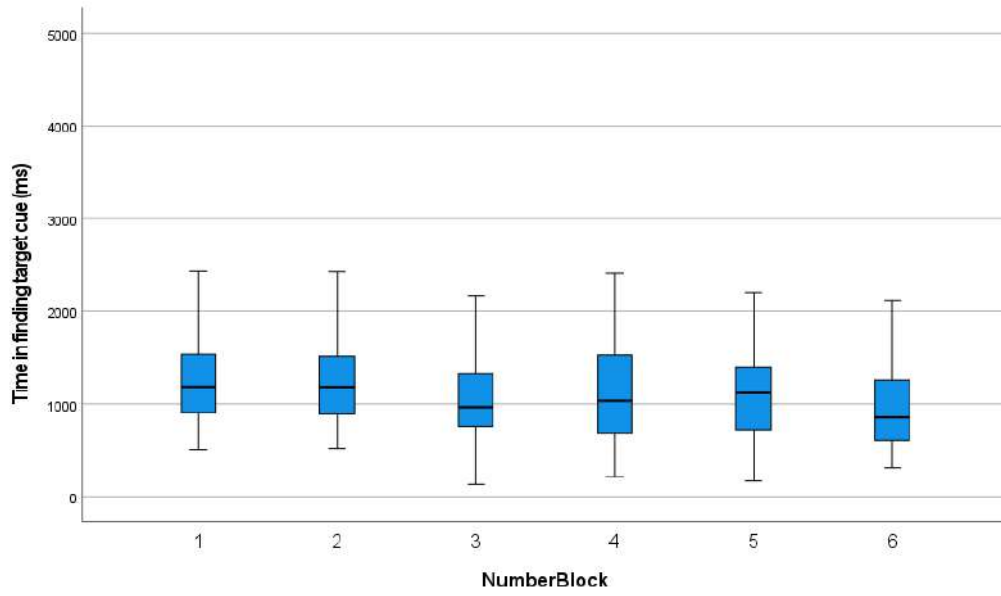
Figure 5.5: Box-plot of each block at the Front target position
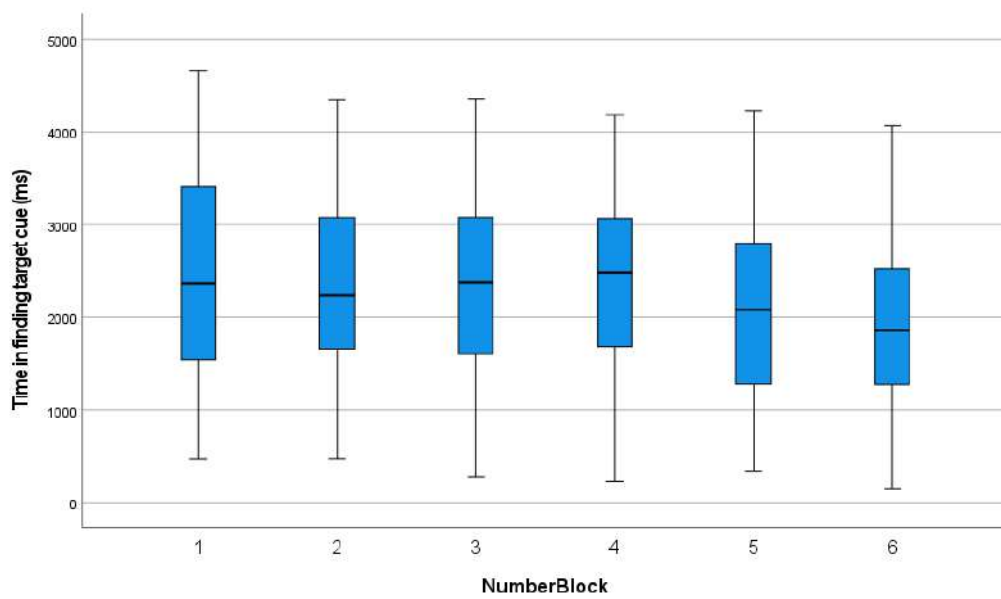
Figure 5.6: Box-plot of each block at the Back target position
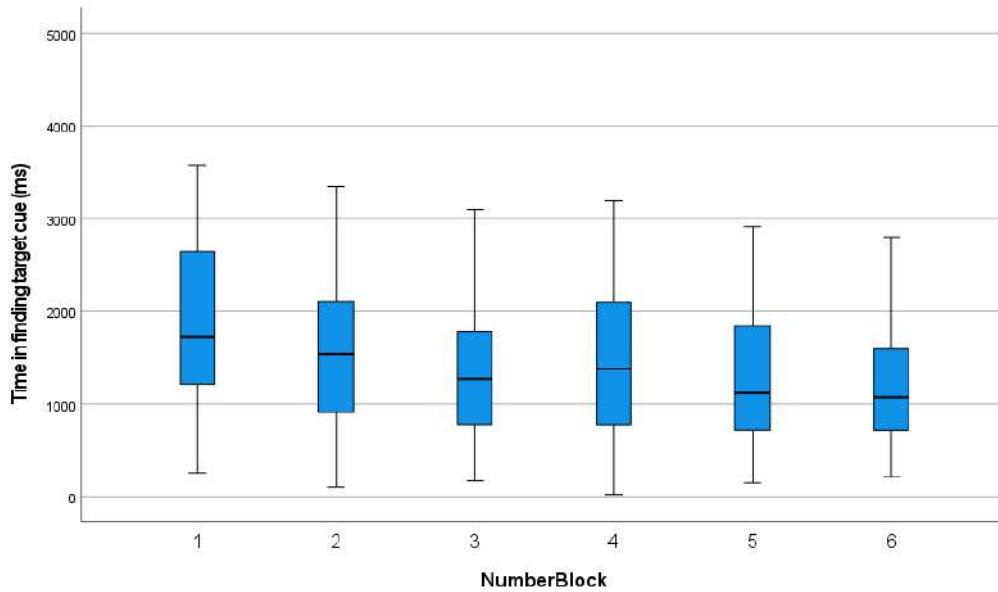
Figure 5.7: Box-plot of each block at the Left target position
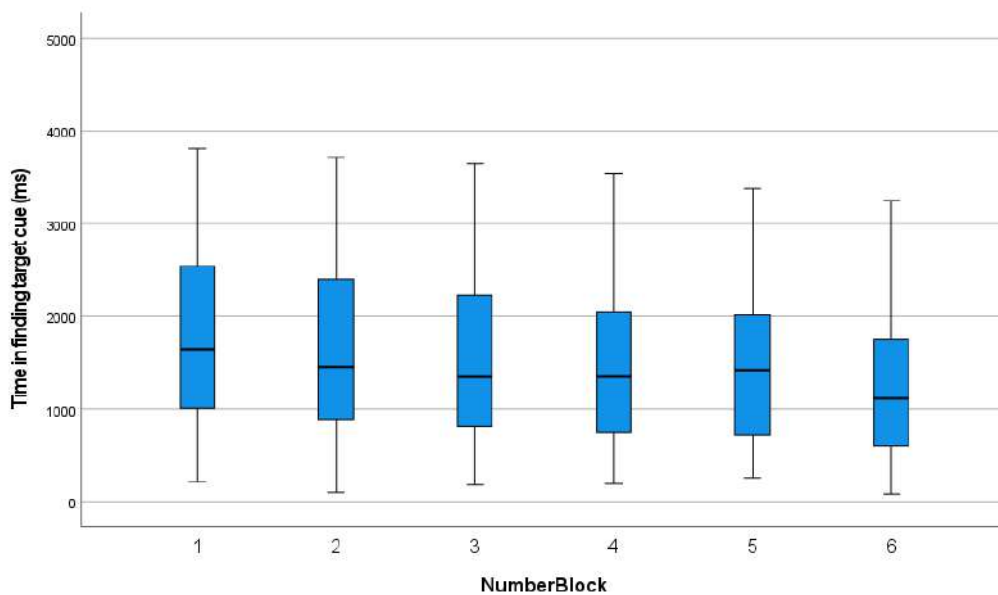
Figure 5.8: Box-plot of each block at the Right target position

From the summary of the linear regression done and the following graph with the box plots, it can be seen that the time in finding the visual cues comes down during the length of the experiment. The only exception in this statement would be the frontal position, which also goes down, but it does more discretely with an Unstandardized B (Target Position) of -54.129. At the same time, the other directions are minus -100 in their Unstandardized B (Target Position), which would mean that there is a bigger learning experience with the front and sides positions. This can be since participants are seated from at the start in the frontal position, and it is thus more comfortable for them to detect the avatar in the front when it is clapping. This can also be seen in figure 4.5, where the values from block 1 are more or less consistent throughout the experiment, meaning that there is not much of a learning experience.

What should be expected when doing an experiment such as this one would be to have results similar on both sides, and the results obtained are close in value. Both values that define the regression line are pretty close, which would help to make a case that the result is similar. But as the variation is so high, it may be with more participants that the values won't be as variate; if more people are put to the test, more conclusions toward this topic could be made. When designing the experiment and discussing the result, the more exciting position was the back one. This is due to the context of the experiment, which was done in a classic classroom chair, in which participants reporter it was uncomfortable to turn to see on the backside. Indeed, the Unstandardized B (Constant) is at its highest with a value of 2617.947 ms, which was expected since, as explained, it usually is the last place the participants would look into. There is one exception, been given by the feedback of several participants. If the previous cue were on one of the sides, the first position to look for the cue would be the same side and then it would go to the sides, so the participant would look in the front and back position. Another aspect is that the back position is not the one that has the lowest Unstandardized B (Target Position) and, therefore, highest learning experience. This may be due to two main factors; the first is that participants throughout the experiment tend to look constantly at the other positions first since there is this sense of discomfort to the back position. The second factor is related to also this sense of discomfort but in another way. When participants hit the trigger indicating they detected the avatar moving, the target and distractor stimuli play out, and they need to decide which type of animal it is. With the help of participants' feedback and further analysis on the head tracking, it has been reported by the participants, that when detecting that the avatar moving is on the backside, get to a more comfortable position to press the trigger button.

5.3 Analysis of the Head Tracking Data

As it has been seen in the analysis of the time in finding the visual cue, there is an interest in studying the head tracking data, to know the strategy that the participants employed during the experiment and to be able to draw conclusions on the degree of head movement.

As stated in another section in the thesis, the parameter analysed in the head tracking data is the yaw rotation. This is the parameter saved in all the experiments, how the participants were turning in the horizontal plane. It is also essential to note that the variable is in Euler angles and that the frontal position corresponds to 90°. Last but not least the sign in the quaternion variable was not taken into account since it was assumed few differences would exist between sides. Therefore the results from both left and right will be of similar value.

In the data-set studied, there is the head tracking data of both when the avatars start clapping and when the participants press the trigger button. Since in this experiment, what is important is the strategy of the participants when looking at the cue, only the head tracking data of when the participants press the button has been analysed.

So, after separating the data between each target position and having checked the assumptions of linear regression in subsections 4.1.1, 4.1.2 and 4.1.3, a linear regression model was done on the data. In the following table, the different values of the importance of this model are shown. Also, the mean and the standard deviation of each position is included, which gives a better picture of how the data is distributed.

Linear Regression Data	Target Position			
	Front	Back	Left	Right
Mean(º)	75.207	64.631	52.675	48.974
Standard Deviation(º)	19.204	23.214	31.693	31.543
Unstandardized B (Constant) (º)	69.886	61.632	59.979	58.569
Unstandardized B (Target Position) (º)	1.548	0.857	-2.087	-2.741
R	0.139	0.063	0.113	0.149
R^2	0.019	0.004	0.013	0.022
Adjusted R^2	0.018	0.003	0.011	0.021

Table 5.3: Linear Regression Data of the Head Tracking data

First, both front direction and sides direction (if looking at the mean value) results show that the cue is in the field of view of the participant. Whereas for the back position, a mean of 64.631 was found with a standard deviation of 23° indicating that the participants were only slightly turned when noticing the back cue.

Again, the values of R^2 and Adjusted R^2, are quite low so there the linear model is not a fit. Therefore, as for the time in finding the visual cue, the linear regression plots can be found in the Appendix (Figures A.17, A.18, A.19 and A.20).

In this subsection, the four plots (Figures 4.9, 4.10, 4.11 and 4.12) will be presented in the same format as the previous section. An important point to note when looking at the box plots is that outliers more than two standard deviations away from the mean have not been taken out of the data set. This is because the interest of this analysis is to have a better understanding of the strategies when looking for the visual cues, so most data has been kept to be able to have a better understanding of this subject.
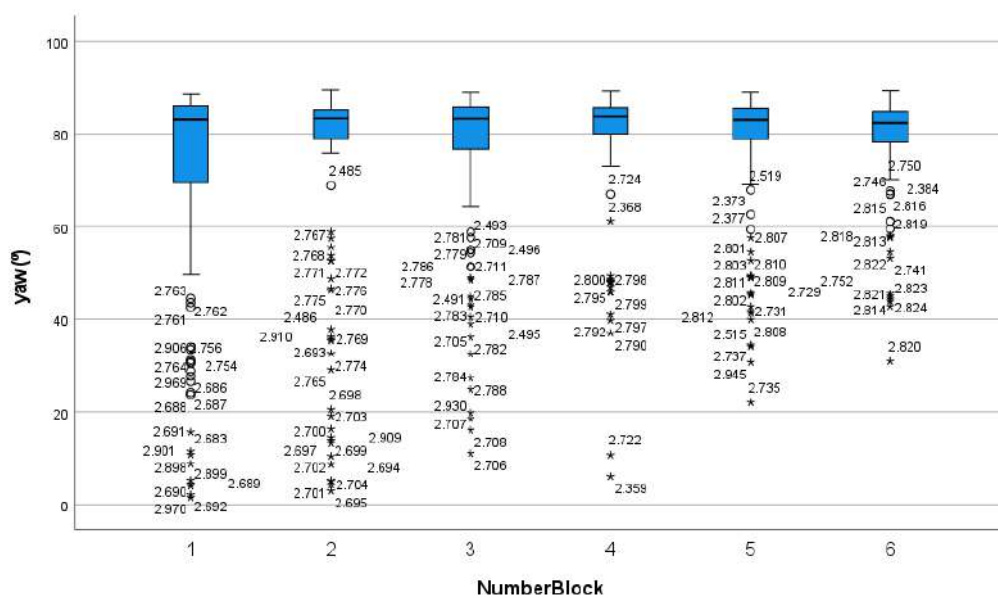
Figure 5.9: Box-plot of each block at the Front target position
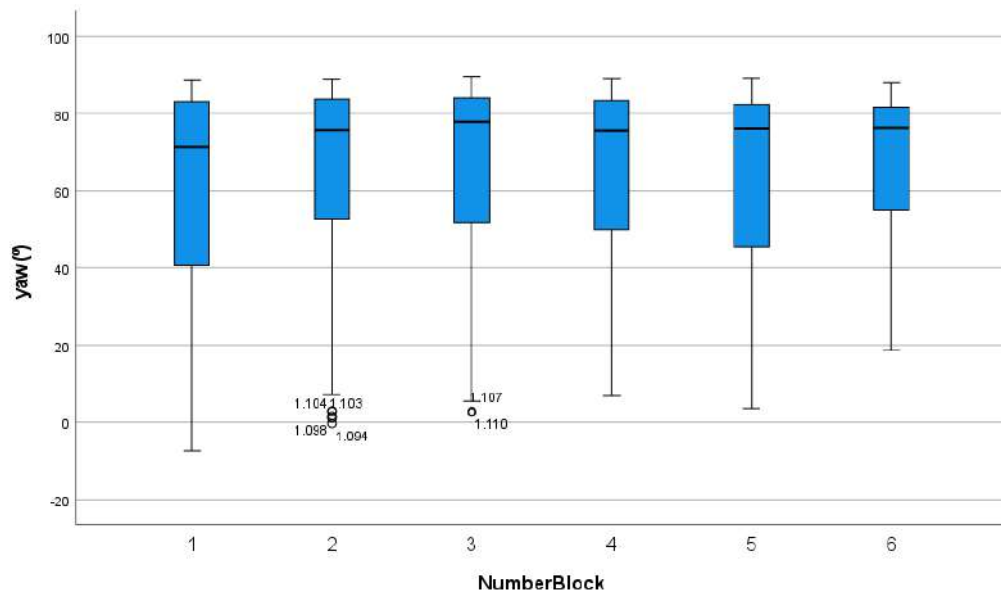
Figure 5.10: Box-plot of each block at the Back target position
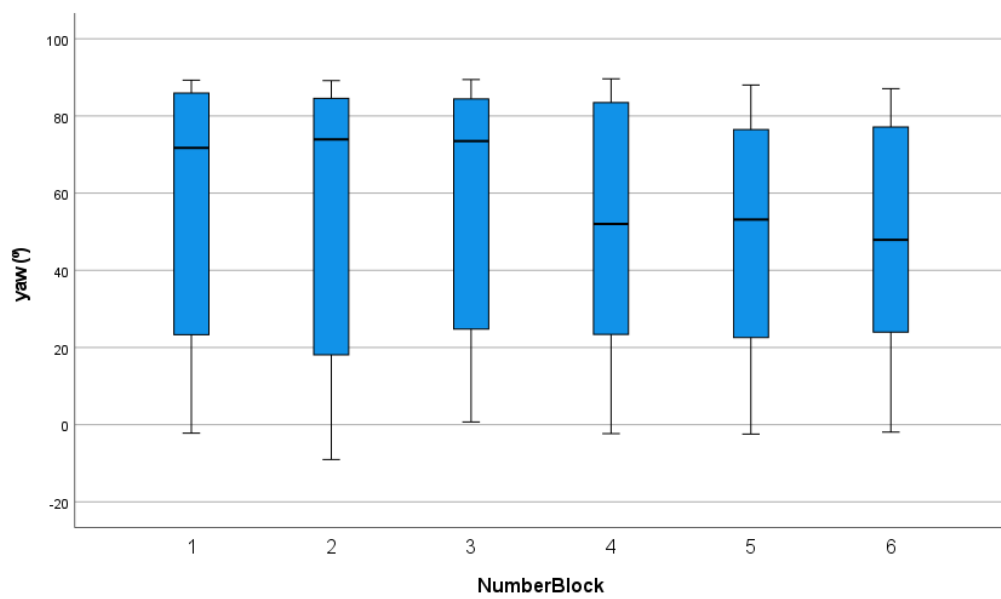
Figure 5.11: Box-plot of each block at the Left target position
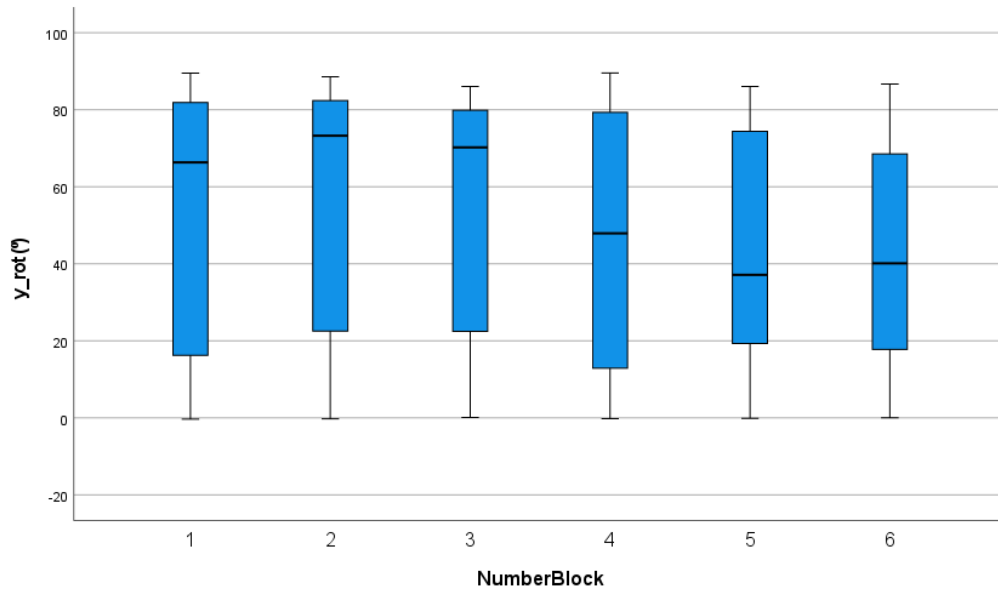
Figure 5.12: Box-plot of each block at the Right target position

Analysing the head tracking data can be seen in both the table with the primary data from the linear model and the box plots. Depending on which position, there is more or less a learning experience throughout the experiment. First and foremost, the front direction, which compared it to the three other positions, is the one that has the most outliers by far. It is also where a more apparent linear model is presented, as participants tend to focus more closely to the 90º mark. This can also be derived from the growing understanding of the experiment as it goes on. Also, since with the Vive headset, the field of view is 110º, most participants had to be at 55º from the cue to see it. Hence, between the range of 35º to 145º for the frontal positions. But since the sign of the quaternion is not taken into account the range would be from 35º to 90º. Having values outside of this visual range is relevant since, from this, it can be seen that some participants during the first stages of the experiment were entirely lost on the experiment in itself (Figure 4.9). This is of crucial importance since the mean is further down in the other positions, almost no outlier shows up, indicating this problem. Another point that it is seen from figure 4.9 is that the variance between blocks comes down by a good margin. Making it seem that participant did have a learning experience with the front position.

For both sides, logical results come out. Both the left and right directions have all value in which the avatar could be seen with the Vive headset field of view. But the mean stays in values near the 45º mark of either direction. Meaning that the most used strategy was to be in a position in which they could already see the side position. They hadn't to go until the 0º value which would mean facing the cue. Also, it can be seen from the data of the linear regression model that on both sides, the rotation was moving closer to the cue as the experiment went on. This is an important discovery because it would hint that participants get more comfortable during the experiment to listen to the stimulus in a more rotated position comparing it to the frontal one. Also the left and right conditions there is a trend of the mean going down by more than 20° indicating that participants probably moved less. However, due to the high variance in the data no real conclusion can be drawn. The Back position is maybe the one that has the most unexpected results. Not only the mean of the rotation is closer to the frontal position than the one to the sides but also the rotation tends to come closer to the frontal position as the

experiment goes on. This could be attributed to two ideas that were given in feedback after the experiment, which would explain this shift compared to other positions. The first one, and already commented, looking at the visual cue at the back is uncomfortable. Several participants indicated that after seeing the avatar clapping on the back, they would return to a position where they felt comfortable before pressing any trigger. This idea would make sense since the data shown in the box plot keeps getting closer to the 90º mark as the blocks go on. The second idea given in oral informal feedback by some other participants was that they didn't need to check whether the back avatar was clapping. As long as they made sure that no one of the other three avatars was clapping, they were convinced that it was the back one that was clapping, and they could then press the trigger in a more comfortable position.

5.4 Other findings

When analysing the data and reviewing the feedback given by all 10 participants, the results of two participants stood out during the analysis (Proband ID=3 and Proband ID=4) which was then contrasted with the informal oral feedback. In this section, these two specific cases will be discussed together with a rough analysis of the data of the other Probands without taking these two into account.

5.4.1 Cybersickness

When designing this experiment, one of the most significant doubts was if the experiment design would give cybersickness to the participants that did the experiment. When the experiment with all the 10 participants was done, some mentioned they had a bit of cybersickness, but participants with Proband ID=4 stated that the last blocks were quite tricky because of the sickness the participant had. While analysing the data and extracting outliers, Proband ID=4 is the participant of which more outliers have been taken out; only 157 of 288 trials have been analysed in the model. Thus, in subsection 4.4.3, a linear regression model has been done without this participant's data.

5.4.2 Different strategy

Lineal Regression Data	Target Position			
	Front	Back	Left	Right
Mean (º)	74.063	73.187	72.190	74.078
Standard Deviation (º)	20.361	22.764	20.792	19.088
Unstandardized B (Constant) (º)	59.981	52.738	54.137	56.898
Unstandardized B (Target Position) (º)	4.024	5.843	5.158	4.909
R	0.340	0.441	0.427	0.442
R^2	0.115	0.195	0.182	0.196
Adjusted R^2	0.103	0.183	0.170	0.184

Table 5.4: Lineal Regression Data of Participant 4

In this table 4.4.2 it can be seen the model for only the participant 4. It can be seen a vast difference from the one extracted for all the participants. First and most noticeable, all the means of all the directions are within 2º, which means that the participant tended to return to a certain pose to listen to the stimulus. What's is more, the Unstandardized B (Target Position) is bigger than four in each of the positions, meaning there is a learning pattern from the participant to get closer to the frontal position as the experiment goes on. Indeed, this learning capability can be seen in the linear regression model key data, which has higher Adjusted R^2 values. Another analysis will be done in the following subsection with the eight Probands remaining.

5.4.3 Revised Results

In this following table, a linear regression model has been done to the data of the eight participants that presented a similar range of values, without participants 3 and 4.

Lineal Regression Data	*Target Position*			
	Front	Back	Left	Right
Mean	74.238	61.042	46.291	41.198
Standard Deviation	19.508	23.137	31.305	30.683
Unstandardized B (Constant)	69.696	52.738	57.160	55.588
Unstandardized B (Target Position)	1.298	5.843	-3.105	-3.906
R	0.114	0.026	0.170	0.218
R^2	0.013	0.001	0.029	0.047
Adjusted R^2	0.011	0.001	0.027	0.046

Table 5.5: Lineal Regression Data 4

From this analysis of only the eight Probands with a similar range of values, two key conclusions can be made. First, as seen on the back position, there is a lesser regression model as seen from the R compared to the other positions where there is an increase. Furthermore, the mean of the back position stay farther away from the frontal position but still gets closer to it as the experiment goes on. The second key point extracted from this last linear regression model is that both sides are closer to the 45º mark, representing the middle point between the cue of the side and the frontal one.

6 Conclusion and Outlook

The thesis aimed to investigate new, more realistic visual cues for an auditory selective attention paradigm in virtual reality. To accomplish it, some research on similar topics was done initially (see section 2.2). One new visual cue was implemented, which required the participant to search for it. The aim was to have a more realistic paradigm in which participants needed to move. To investigate the influence of this new cue on the overall paradigm, the goal was to see how people would search for it, how much time they would take to search for the cue, and what their head tracking data would be when searching for it.

Later on, a listening experiment was done with 10 participants with this new experiment with the new visual cue implemented. Before the experiment, it was made sure that all participants had good visual and hearing levels. Training in the VR was done before starting the experiment. The data of the experiment was then recorded and also informal oral feedback was given when the experiment was done.

The data was then analysed with linear regression models since it was believed that participants would change their strategy when finding the cues as the experiment went on. What was seen was that there was a tendency for the time it took to find the cue decreased throughout the experiment, showing some learning experience in the VR experiment. Also, with the head tracking data, for all the positions but the back one, it was seen that people didn't listen to the stimulus near the position in a majority, though they tended towards it as the experiment went on. The exception to this claim would be the back position in which clues would indicate that strategy from participants would be to look at the cue and wait to be in a more comfortable position to listen to the stimulus. It would also indicate in some cases that participants, when they made sure that any other avatar was clapping, would assume that the back one would and wouldn't turn to make sure.

Further analysis coupled with the informal feedback also showed that one out of ten participants suffered cybersickness, and its results weren't the most relevant. Also, another one of the participants showed very different results from the other participants since they stayed focused in a more comfortable position close to the frontal one. An analysis was then done with the eight participants remaining, and it was seen that the average of the head tracking data for the sides stayed closer to the 45º mark respecting the frontal direction, which would mean that participants would hear the stimulus in a position in which they can have in check at least two of the positions.

For further research, it would be recommended to see whether this shift of strategy between one of the participants was a case outside of the norm or do some people act as this participant did. With much more extensive participation in this experiment or a similar one, it could be seen how much this case is relevant for this research. It is recommended for further research to use this cue, because with the informal feedback it was seen that most participants didn't feel sick

with the experiment, and there is many more variables that could be analysed. This thesis only focused in the head tracking when pressing the trigger but also the head tracking at all time would be certainly a potential further research

A Appendix
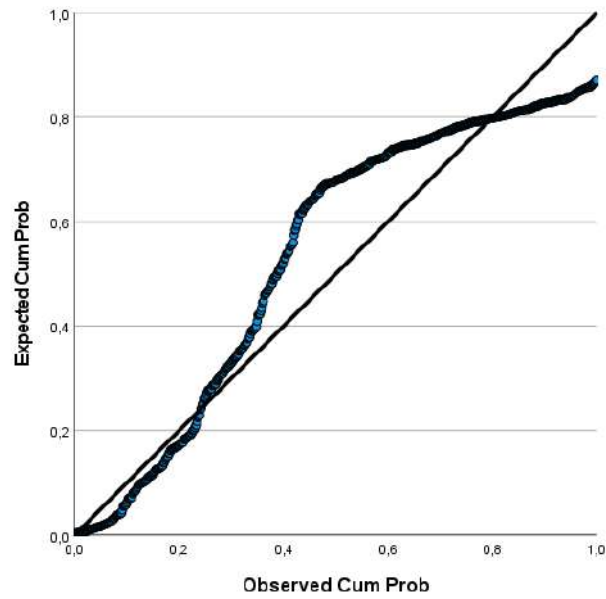
Figure A.1: P-P plot of the head tracking data when the Target Position is the back one
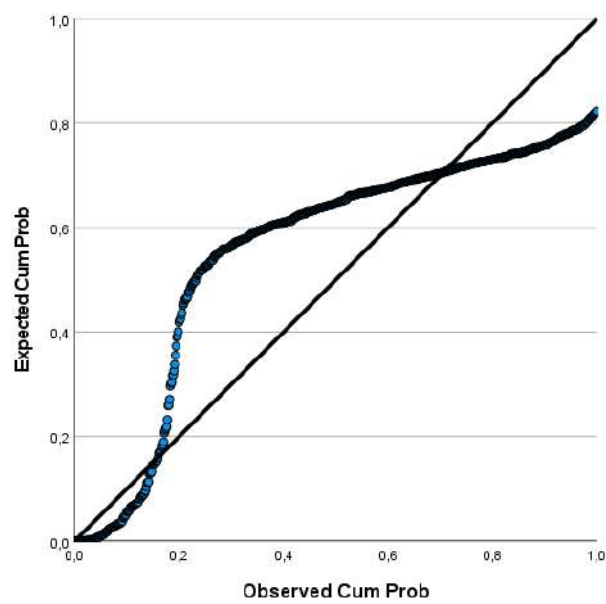
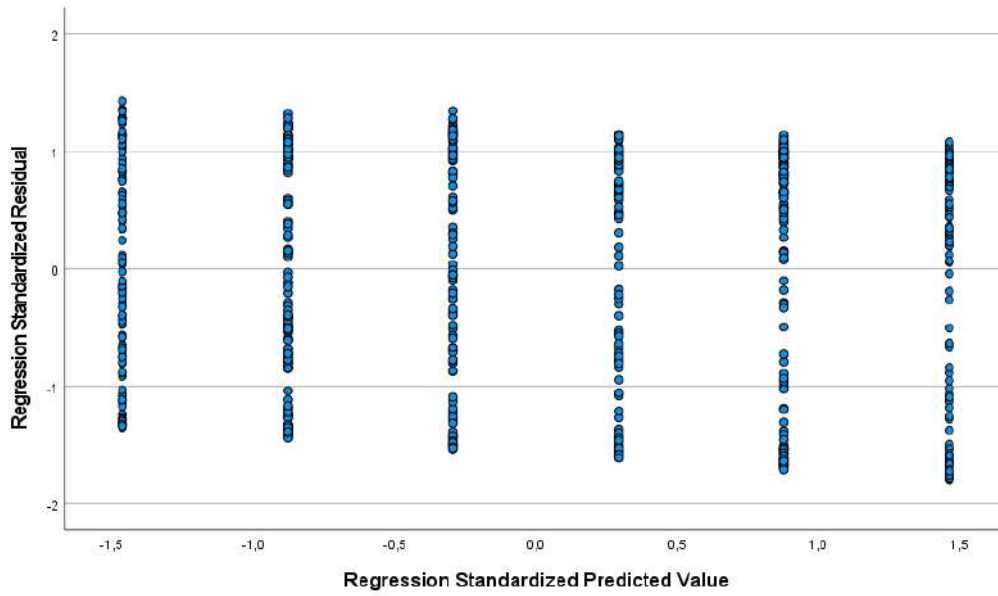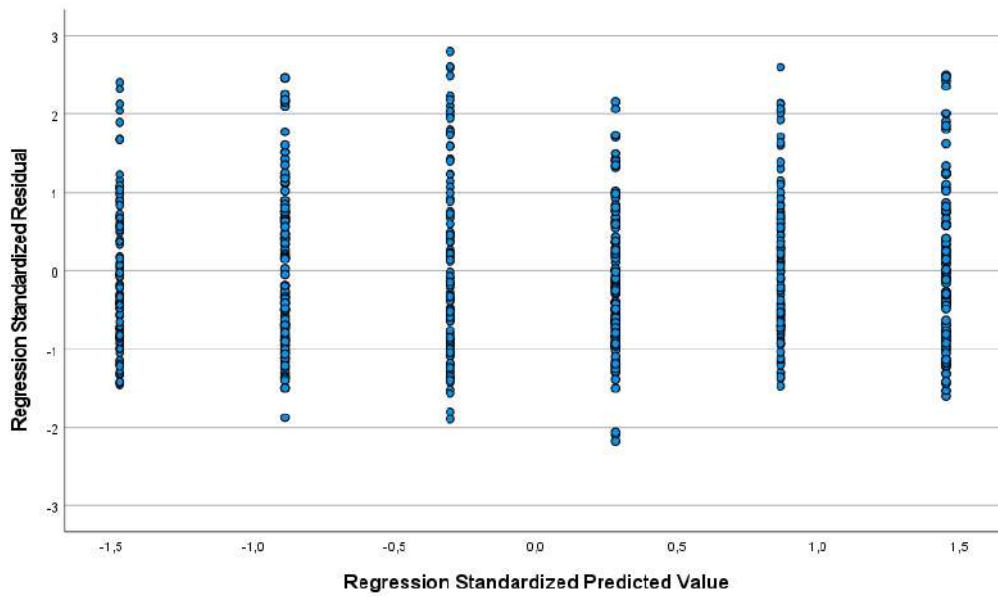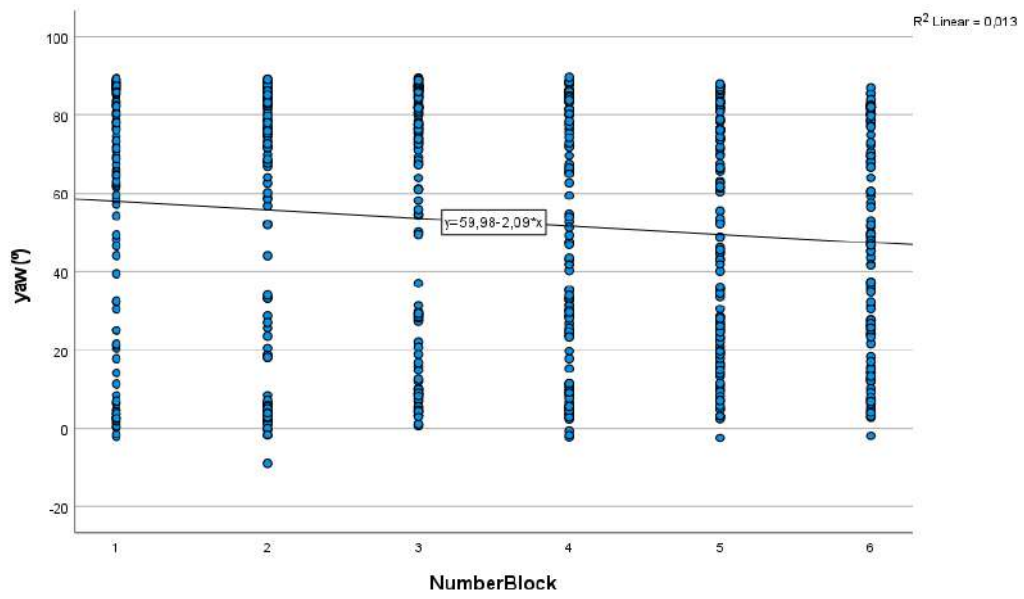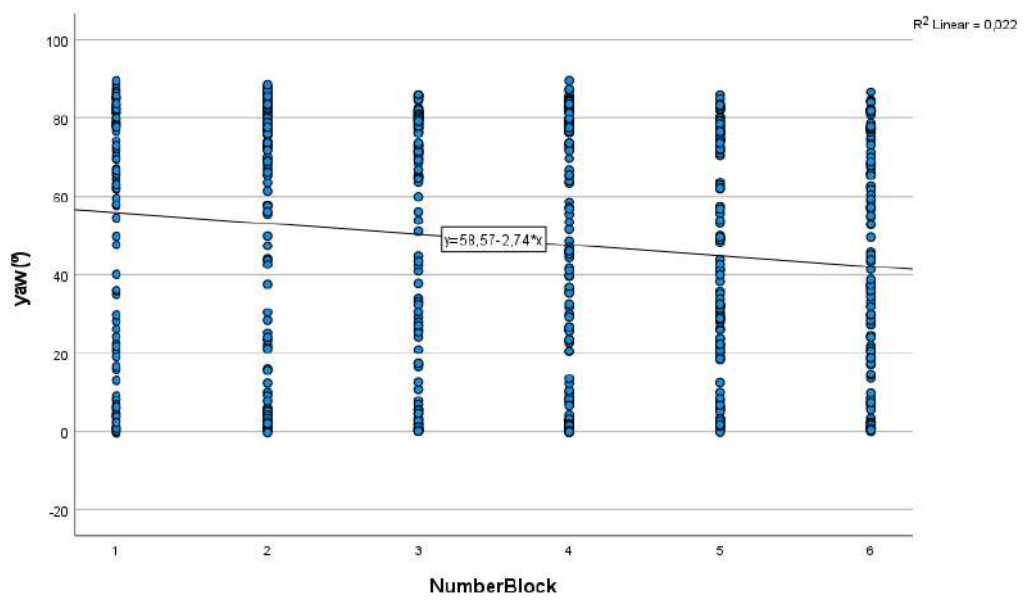Figure A.2: P-P plot of the head tracking data when the Target Position is the front one
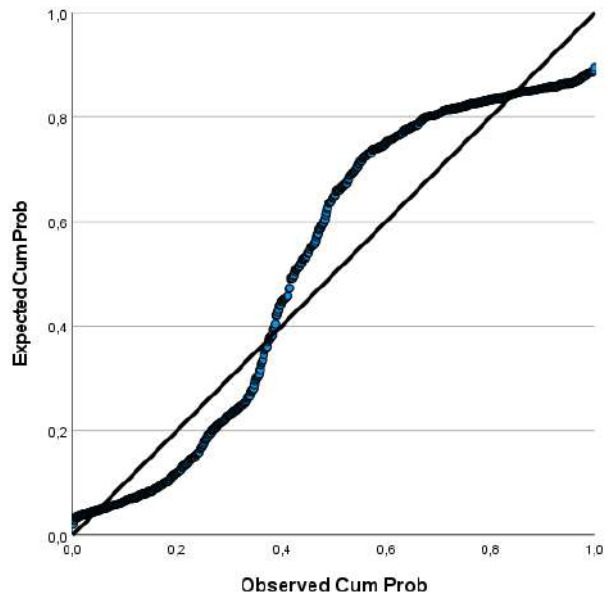
Figure A.3: P-P plot of the head tracking data when the Target Position is the left one
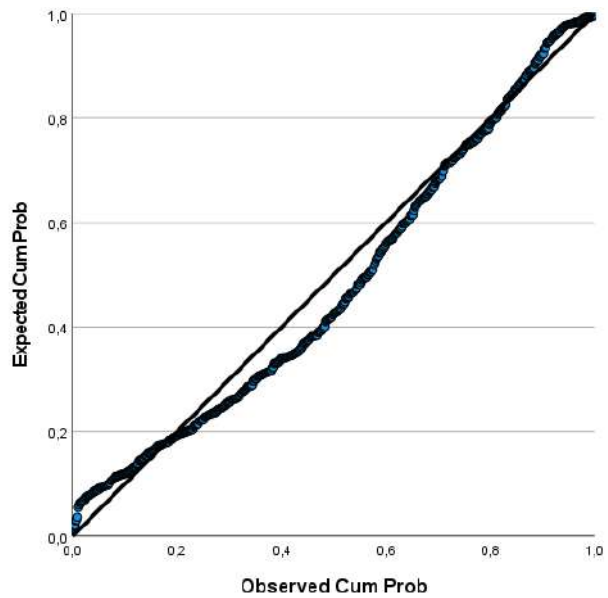
Figure A.4: P-P plot of the time finding the visual cue data when the Target Position is the front one
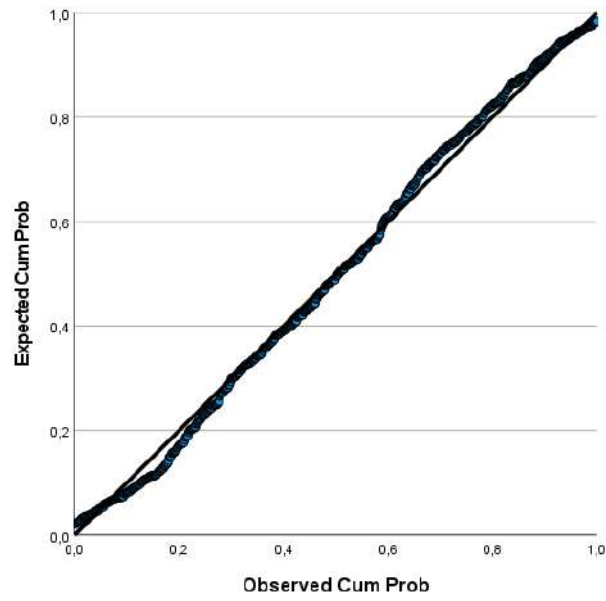
Figure A.5: P-P plot of the time finding the visual cue data when the Target Position is the back one
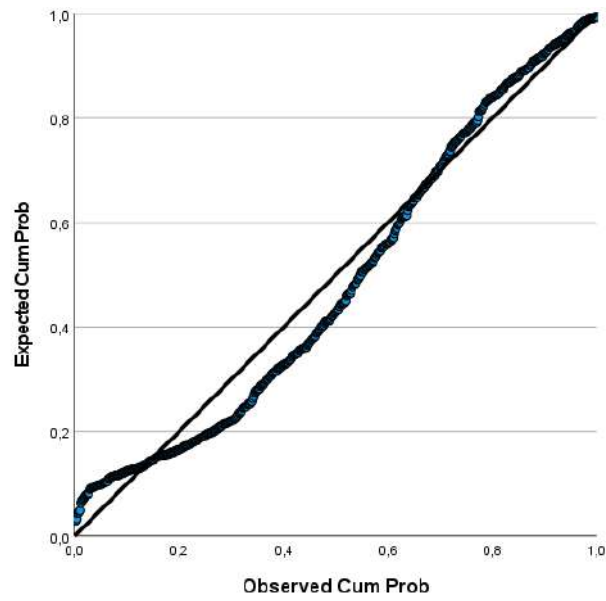
Figure A.6: P-P plot of the time finding the visual cue data when the Target Position is the right one
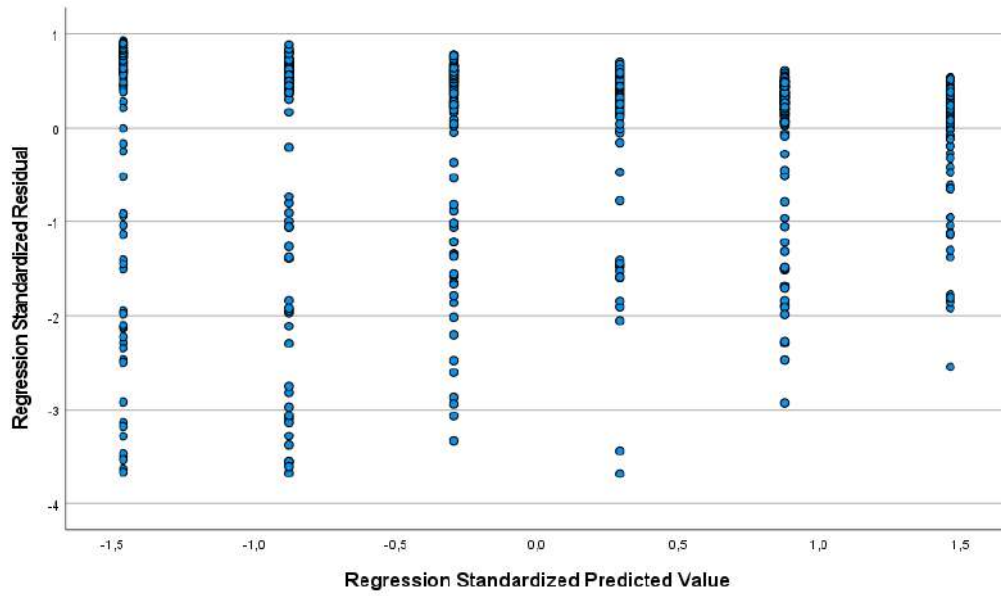
Figure A.7: Scatter plot of residuals of the head tracking data when the Target Position is the front one
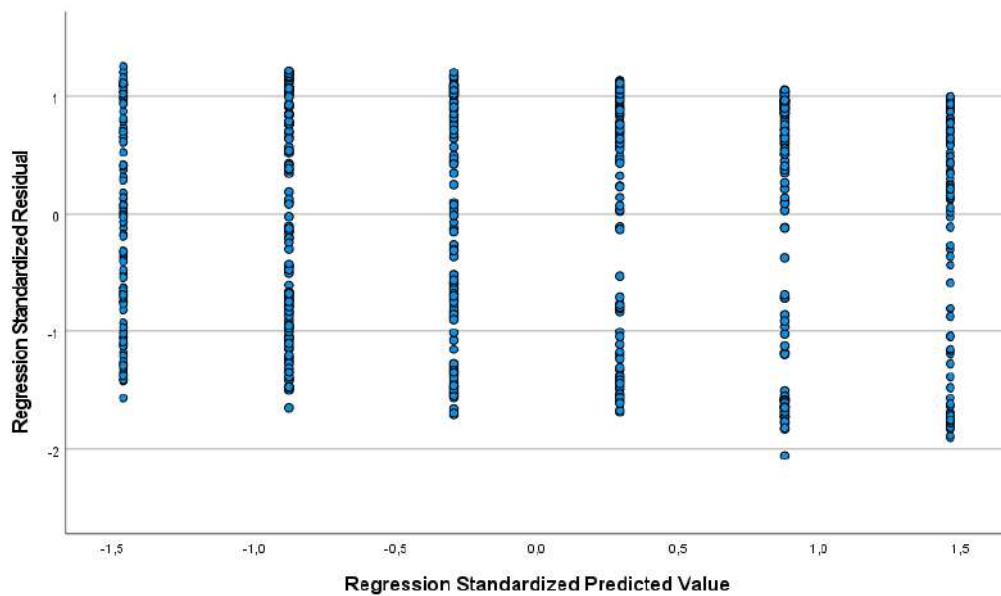
Figure A.8: Scatter plot of residuals of the head tracking data when the Target Position is the left one

Figure A.9: Scatter plot of residuals of the head tracking data when the Target Position is the right one

Figure A.10: Scatter plot of residuals of the time finding the visual cue data when the Target Position is the front one
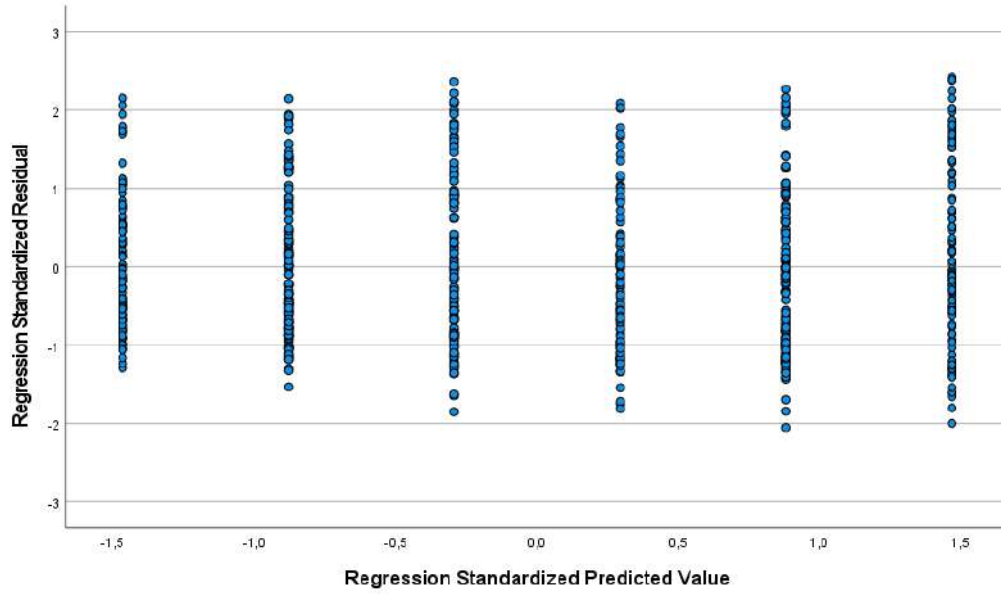
Figure A.11: Scatter plot of residuals of the time finding the visual cue data when the Target Position is the left one
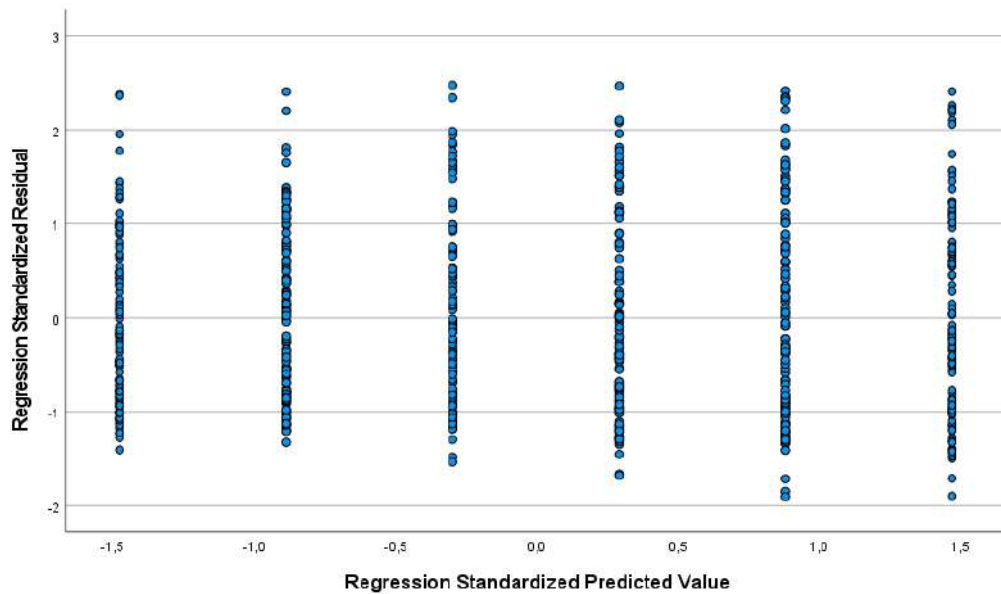
Figure A.12: Scatter plot of residuals of the time finding the visual cue data when the Target Position is the right one
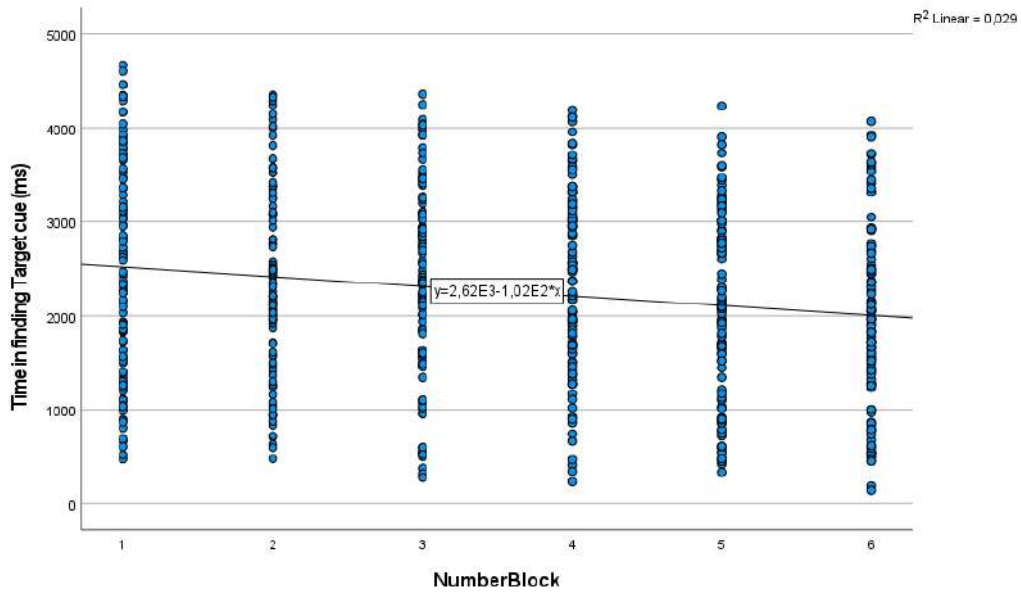
Figure A.13: Linear Regression of the time in finding the target cue data when the Target Position is the back one
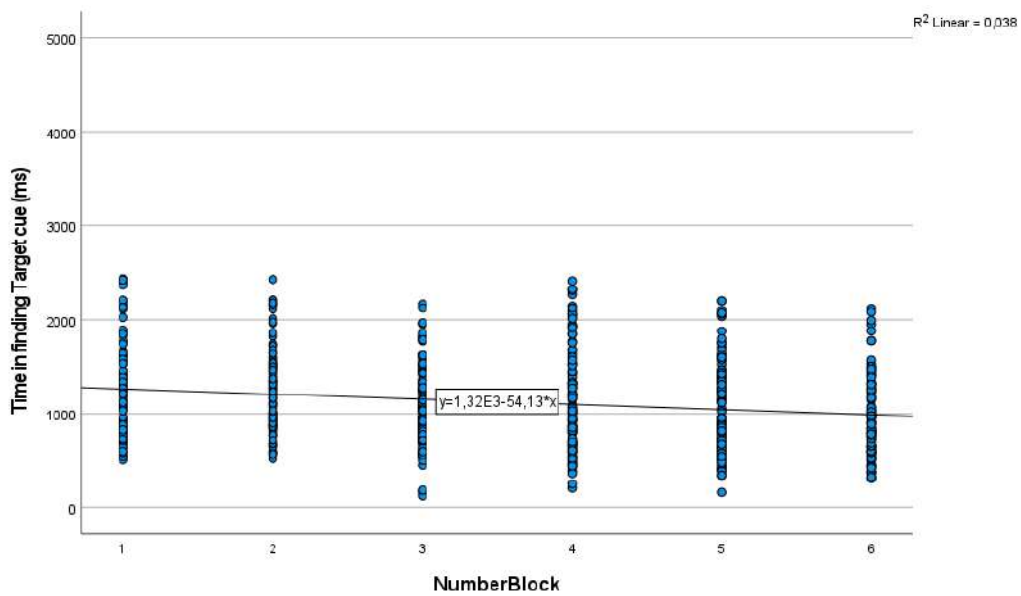
Figure A.14: Linear Regression of the time in finding the target cue data when the Target Position is the front one
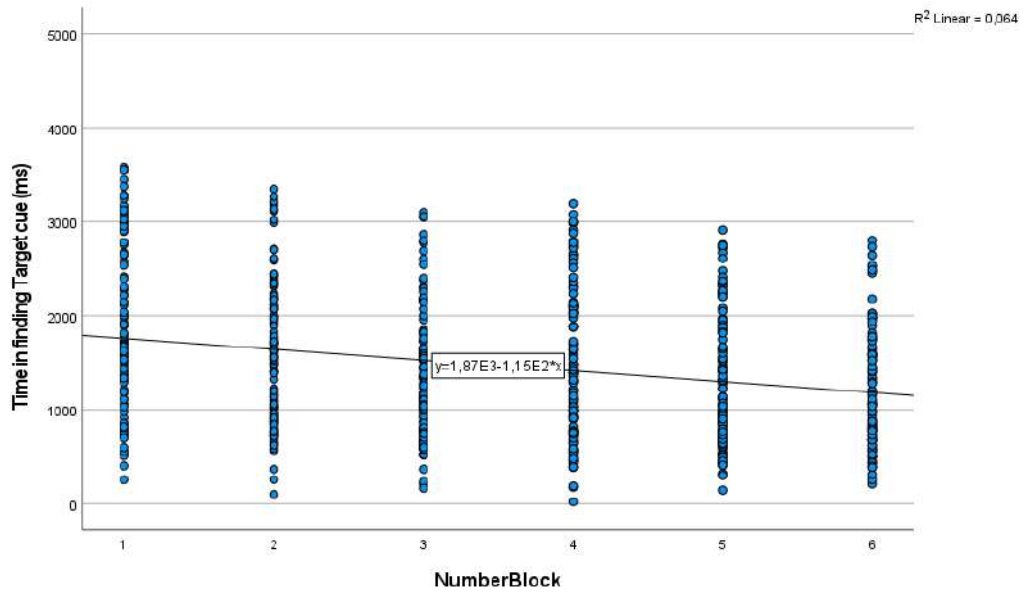
Figure A.15: Linear Regression of the time in finding the target cue data when the Target Position is the left one
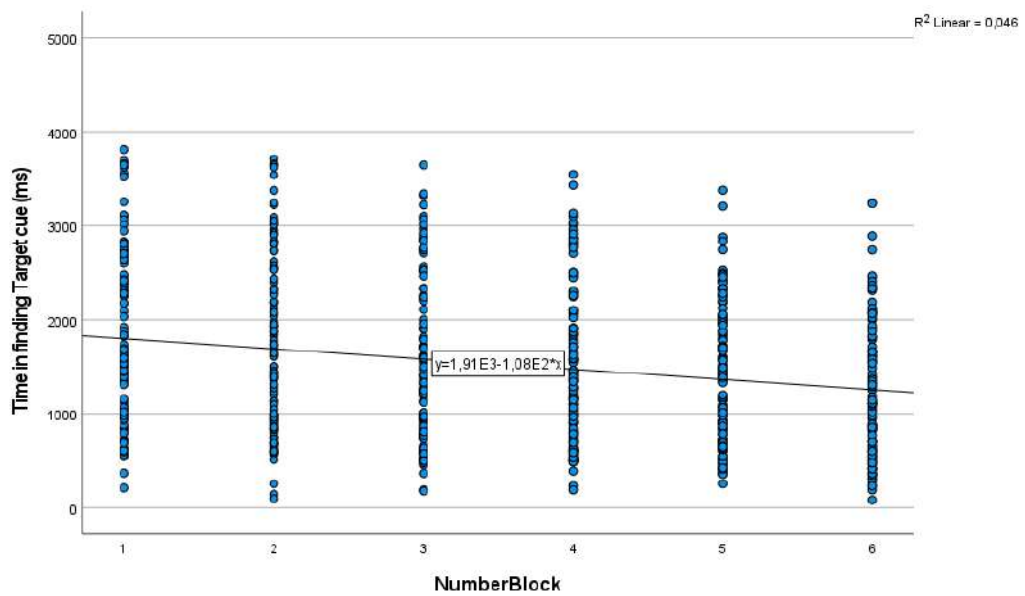
Figure A.16: Linear Regression of the time in finding the target cue data when the Target Position is the right one
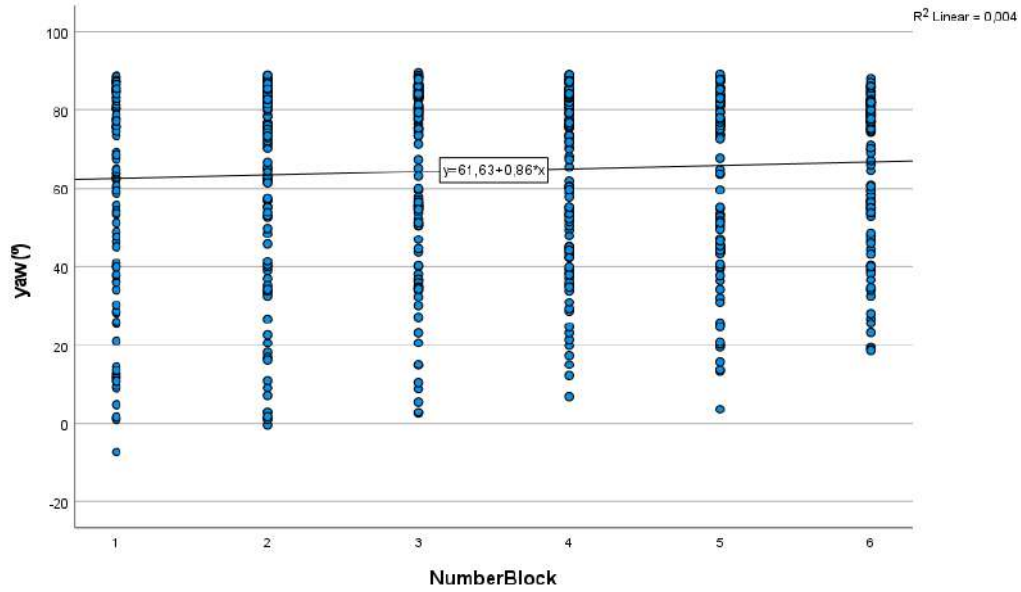
Figure A.17: Linear Regression of the head tracking data when the Target Position is the back one
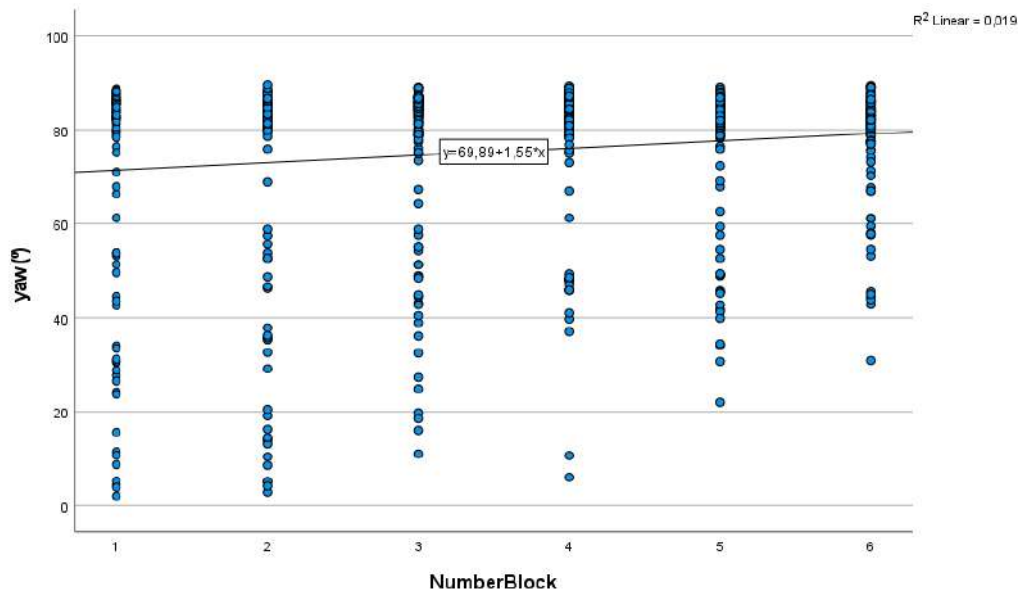
Figure A.18: Linear Regression of the head tracking data when the Target Position is the front one

Figure A.19: Linear Regression of the head tracking data when the Target Position is the left one

Figure A.20: Linear Regression of the head tracking data when the Target Position is the right one

List of Figures

List of Tables

Bibliography

[1] ADOBE SYSTEMS INCORPORATED: Mixamo. `https://www.mixamo.com/#/`, . – Online; accessed 30 March 2022

[2] APA DICTIONARY OF PSYCHOLOGY: Head-related transfer function (HRTF). `https://dictionary.apa.org/head-related-transfer-function`, . – Online; accessed 31 March 2022

[3] BLENDER ONLINE COMMUNITY: Blender - a 3D modelling and rendering package. `http://www.blender.org`, . – Online; accessed 30 March 2022

[4] CHERRY, Colin: Some experiments on the recognition of speech, with one and with two ears. In: The Journal of the acoustical society of America 25 (1953), S. 975–979

[5] DIETRICH, Pascal ; GUSKI, Martin ; POLLOW, Martin ; MÜLLER-TRAPET, Markus ; MASIERO, Bruno ; SCHARRER, Roman ; VORLÄNDER, Michael: ITA-Toolbox – An Open Source MATLAB Toolbox for Acousticians. (2012)

[6] IBERDROLA: Virtual Reality: another world within sight. `https://www.iberdrola.com/innovation/virtual-reality`, . – Online; accessed 31 March 2022

[7] JOHNSSON, Roger ; ODELIUS, Johan ; NYKÄNEN, Arne: Self assessment of speech intelligibility listening to binaural recordings. (2009)

[8] KOH, Iring ; LAWO, Vera ; FELS, Janina ; VORLÄNDER, Michael: Switching in the Cocktail Party: Exploring Intentional Control of Auditory Selective Attention. In: Journal of Experimental Psychology: Human Perception and Performance, American Psychological Association 37 (2011), Nr. 4, S. 1140–1147

[9] LOH, Karin ; FINTOR, Edina ; NOLDEN, Sophie ; FELS, Janina: Comparing intentional switching of auditory selective attention in children and adults in an experiment suited for children. (2019)

[10] LOH, Karin ; FINTOR, Edina ; NOLDEN, Sophie ; FELS, Janina: Children's intentional switching of auditory selective attention in spatial and noisy acoustic environments in comparison to adults. (2020)

[11] MASIERO, Bruno ; FELS, Janina: Perceptually Robust Headphone Equalization for Binaural Reproduction. (2011)

[12] OBEREM, Josefa ; LAWO, Vera ; KOH, Iring ; FELS, Janina: Intentional Switching in Auditory Selective Attention: Exploring Different Binaural Reproduction Methods in an Anechoic Chamber. In: Acta Acustica United With Acustica 100 (2014), S. 1139–1148

[13] REIMERS, Carolin ; LOH, Karin ; LEIST, Larissa ; FREMEREY, Stephan ; RAAKE,

Alexander ; Klatte, Maria ; Fels, Janina: Examining Auditory Selective Attention in a Classroom Setting using Audiovisual Virtual Reality. (2021)

[14] Statistics Solutions: Assumptions of Linear Regression. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/assumptions-of-linear-regression/, . – Online; accessed 31 March 2022

[15] Statistics Solutions: Homoscedasticity. https://www.statisticssolutions.com/free-resources/directory-of-statistical-analyses/homoscedasticity/, . – Online; accessed 31 March 2022

[16] The MakeHuman team: MakeHuman 3D Model Creation. http://www.makehumancommunity.org/, . – Online; accessed 30 March 2022

[17] tobiipro: How fast is human perception? https://www.tobiipro.com/learn-and-support/learn/eye-tracking-essentials/how-fast-is-human-perception/, . – Online; accessed 31 March 2022

[18] Willert, Volker ; Eggert, Julian ; Adamy, Jürgen ; Stahl, Raphael ; Körner, Edgar: A Probabilistic Model for Binaural Sound Localization. (2006)

Eidesstattliche Erklärung

Hiermit erkläre ich, dass ich die vorliegende Arbeit selbstständig und ausschließlich unter Verwendung der angegebenen Hilfsmittel erstellt habe, sowie dass wörtliche oder inhaltliche Zitate entsprechend kenntlich gemacht und dem Literaturverzeichnis zu entnehmen sind.

Aachen, den April 29, 2022

de Lapparent Alvarez	Pol	453432
Name	Vorname	Matrikelnummer

Unterschrift