UNIVERSITÀ DEGLI STUDI DI PADOVA

DIPARTIMENTO DI INGEGNERIA DELL'INFORMAZIONE
LAUREA MAGISTRALE IN INGEGNERIA INFORMATICA

# A VIRTUAL REALITY ENVIRONMENT WITH PERSONALIZED SPATIAL AUDIO RENDERING

*Laureando:*
Fabio Prandoni

*Matricola:*
1109431

*Relatore:*
Proff. Federico Avanzini

*Correlatore:*
Dtt. Michele Geronazzo

Anno Accademico 2016/2017
Data Laurea 10/04/2017

# Abstract/Summary

This thesis is about enhancing audio perception in virtual reality environments taking advantage of a developed matlab tool for the extraction of HRTFs (see section 1.4.1). The developed matlab application can be used also by unskilled people and, after a simple installation on a personal computer, it is ready to be run. When the selection of the best HRTFs set for a subject is completed, it can be possible to add that information to virtual reality environments. In particular, tests were done in an android unity based virtual reality environment with a Samsung Gear + Samsung S7 head-set, taking measurements about perception of sounds with different HRTFs conditions for different people.

In chapter 1, the main concepts of virtual reality, VAD systems and 3D audio rendering and spatial hearing are discussed. In chapter 2, the matlab application is presented with a guidline of how to use it in addition to the physics theory behind it. In chapter 3, additional tools for the experiment are discussed and illustrated. In the last chapter (4), tests and results for different subjects are presented.

# Ringraziamenti

Vorrei scrivere qualche grazie sentito per questo traguardo che ho raggiunto.

Dicono che abbia il dono della sintesi. Vedro' di farci star tutto in breve scrivendo una riga super riassuntiva per ognuno. Un'altra tesi non la scrivo, anche se probabilmente servirebbe.

### Family

In particolare, un grazie va ai *miei genitori* che mi hanno sempre sostenuto, a modo loro per carita', nel bene e nel male il piu' che potevano.

Al mio *Papalino* che ha dimostrato che e' sempre presente nei momenti di bisogno e che mi ha sempre spronato a dar di piu', anche quando gia' avevo ottimi risultati.

Alla mia *Mamy* che nonostante le incomprensioni, ci mette sempre tutto l'amore che ha, nelle cose che fa per me (correzione dell'inglese della tesi incluso).

Alla mia *Luciu* per il supporto che mi da', veramente di qualsiasi tipo. Sarebbe oggettivamente tutto diverso senza di lei, a casa Prandoni, dalle piccole cose alle grandi.

Al mio *'Cello* per essere stato un esempio concreto da cui prendere spunto per puntare a cio' che piu' ami inseguendo le proprie passioni.

Ai miei quattro nonni *Enrico, Enrica, Remo e Mirella*. Mi hanno aiutato a crescere e sempre viziato per farmi felice con ogni loro mezzo.

Ai miei zii/zie da ogni dove. Che ho imparato a conoscere meglio in questi ultimi anni e a cui voglio bene. Ognuno ha messo la sua fetta per farmi arrivare dove sono come sono. Chi con gli scacchi (*Andrea*), con i massaggi (*Silvano*), con compleanni arabeggianti e solarita' innate (*Elena*), violoncelli e musica (*Marco*), con acconciature coccolose (*Roberta*), palacinche zuccherose (*Carlo*).

Ad ogni sorta di cugino che ho, piccolo e grande. Che mi ha tenuto compagnia sicuramente piu' di una volta tra giochi e divertimenti vari. Sperando di non dimenticarne nessuno *Leone,Anastasia,Giovanni,Giulia, Paolo, Silvia e Guido*. Chissa' anche se *Leonardo e Nicola* faranno parte della tribu'.

**Teaching**

Un grazie speciale va ai miei professori e maestri di ogni scuola che ho frequentato. Ognuno ha saputo insegnarmi al meglio cio' che tuttora so. In particolare vorrei ringraziare la mia professoressa di matematica delle superiori Negrini che mi ha dato le basi per continuare con il mio percorso. E un grazie alla mia prima maestra in assoluto Alessandra, che ha dato il via alle danze.

Per parlare di cose piu' recenti, un grazie al professor Federico Avanzini e Michele Geronazzo che mi hanno guidato in questi ultimi mesi. Merita di essere nominato anche Erik che mi ha fornito codice di partenza per i lavori che ho portato a termine ed e' stato sempre gentilissimo per consigli e bug celati da risolvere.

**Amici**

Un grazie anche a tutti gli amici che mi hanno seguito in questi anni ognuno nel suo campo.

Ad *Alessia* che pazientemente sa insegnarmi parole nuove ogni volta, facendo un po' la saccente diciamocelo. Dice che ha corretto il capitolo 3...Se li' trovate errori, e' quindi colpa sua.

Ad *Alex*, mio Bro. Per essermi stato vicino per anni anche a 300km di distanza. Spesso lo devo chiamare io per sentirlo, ma recentemente anche lui ha scoperto il tastino delle chiamate. There is hope.

Ad *Eleonora*, regina del D-minor. Mi ha sopportato per piu' di meta' mio album mettendoci la sua magica voce. Sapra' mai scrivere una canzone in una diversa tonalita'?

A *Giadina* che e' sempre pronta a supportarmi in ogni momento (anche vedendoci e sentendoci veramente poco), partendo dalle elementari. Adesso che si sta dilettando in spettacoli pirotecnici, speriamo non si faccia troppo male.

A *Perucchi* che gia' si stava disperando che dopo tutto sto tram-tram non lo citavo nemmeno per il suo lavoro. E' stato un grande compagno di progetti e di esami. Senza di lui i miei lavori sarebbero stati molto piu' a caso e soprattutto scritti molto piu' a caso. Non potevo chiedere di meglio come collega, dato che si e' trasformato in un ottimo amico. Ottimo smanettone, ma con la matematica formale ha qualche bug.

A *Wash*, mio compagno di Heartstone. Riesce sempre a farmi ridere entro due minuti di conversazione. Nulla da togliere al fatto che nello stesso tempo mi faccia anche imbestialire.

A *Kranius* che mi ha saputo tenere compagnia durante questi due anni in serate speciali. Prima o poi, rigiochera' ad un gioco in cui ha perso mentre gli ho scombato addosso?

A *Gardus Bonus Pastroccius*, compagno di viaggio da tempo immemore (no beh me lo ricordo). Nonche' consulente Telecom. La ringraziamo per i suoi servigi e siamo contenti di aver scelto la sua compagnia. Cordiali saluti. Ci scusiamo per il disagio. A prescindere.

A *Miriam e Stefano* tra i miei piu' cari amici di infanzia. Non penso di aver mai riso con nessuno come con loro. E soprattutto per cose talmente fuori dal mondo che sarebbero da manicomio. Io dico solo...."alle galline!".

A *Raffy* per la compagnia a distanza zero, gridando a gran voce da fuori la strada. Le chiacchiere su strada sono state un buon modo per divertirci e rilassarci devo ammettere.

# Contents

# List of Figures

# List of Tables

# Chapter 1

# Introduction

## 1.1 Virtual Reality

### 1.1.1 What is Virtual Reality?

Virtual Reality (VR) is the match between ”Virtual”, a simulation of something with electronic devices and ”Reality”, the perception of the real world where humans live. Normally, it is implemented by consoles, smartphones and PCs using a visor for the hardware and, for the software, by some game engines that are able to render this kind of technology. In technical terms, it is a way to describe a three-dimensional computer generated environment which can be explored and alterated.

### 1.1.2 Applications

Nowadays VR is used in many fields, from educational purposes to entertainment, connecting people, simulations, prototyping, manufacturing, engineering, fashion, tourism, etc...Some of these application fields have already seen light in markets and common people lives giving the idea of a strong growing future technology.

Below, some famous examples of VR applications are shown, proving that this new technology has the potential to make some radical changes in our habits.

#### Gaming

World-wide gamers had waited for this technology for a very long time. It is a completely new way of playing video games and interact with artificial intelligence, using more than ever human senses in a more similar real-life way with respect to screens. Some PC and console manufacturers have developed some virtual reality games in a very high quality environment with respect to past. One example is Playstation 4 that has on the market games like: *Until Dawn: Rush of Blood*, *Batman: Arkham VR*, *Resident Evil 7 Biohazard* and other titles that lots of people are playing in their consoles in everyday life.

**Figure 1.1:** *Playstation VR and Playstation Move. Picture taken from https://www.theguardian.com*

**Educational**

Several VR applications have already been tested in schools and online for more immersive teaching and learning experiences. Entire sites and apps are dedicated to that purpose and http://immersivevreducation.com is just one of them with different kinds of VR experiences for all students who wanted to live in first person some events like Apollo 11 moon landing or a Titanic tour on the boat in 1912. There are lots of surgery simulations too, ready to make medical preparation stabler, more flexible and possible.

**Art-Fashion**

*Tiltbrush* is a VR 3D-painting application made by Google. It permits the user to make 3D light and colorful painting and drawings surpassing the 2D limitations of paper and pencils. It's available online on Steam for about *30$* for HTC Vive. It could be used to have fun and make the drawing experience a complete body experience as well as to create some fashion prototypes that could be shown in a complete 3D virtual environment.


**Figure 1.2:** *Screenshot from the full Youtube presentation: Tilt Brush - Painting from a new perspective.*

#### Connecting people

In the Oculus Connect 3[1], the third annual virtual reality developer converence, Facebook CEO Mark Zuckerberg created a new standard for connecting people. He stayed on the stage for 8 minutes with two his friends that were at their home long away from that event. In that period of time, they spoke using an avatar that moved hands, had facial expressions and the possibility to do a lot of VR smart activities. They changed the background during the conversation from deep ocean, to Mars land and Facebook offices. Moreover, they played some games, like chess and cards, interacting also with virtual screens and devices that shared that virtual encounter on Facebook as if it were the real world.

### 1.1.3 Improving Audio

Some of the above mentioned applications are not striclty related with audio listening. On the other hand, others are more linked with audio and the perception of sounds. For instance, a VR concert or game scene, needs a more accurate sound localization for high quality fedelity representation. That is why it is important to pay attention to audio rendering and spatialization of sounds. The next sections are about improving audio listening in virtual environments, VADs,

[1]October 5th - 7th, 2016. http://www.texample.net/tikz/resources/


**Figure 1.3:** *Screenshot from the full Youtube presentation: Facebook Social VR Demo - Oculus Connect 2016.*

spatialization of sounds and Head-Related Tranfer Functions (see section 1.4.1 for a formal definition).

## 1.2 Virtual auditory display (VAD)

A Virtual Auditory Display (VAD) is a system for generating sounds with spatial positional characteristics and conveying them to a listener.

A dynamic 3-D audio system (like VR), in which the user movements change the sound environment, requires the following features:

1. **Head Tracking**: one of the major problems for virtual 3-D audio systems is head movement [5]. Because of that, more attention is assigned to head tracking, which is particularly important in 3-D virtual environment; in the 1.2.1 subsection the head movement tracking possible implementations are shown.

2. **Fast Processing**: the tracking algorithm to make changes in virtual sound properly correspond to changes in the listeners position and orientation. Ideally the brain should not detect any delay between change in position/orientation and change in sound properties. That is sure linked also with the device that render the environment regard 3-D audio-video

3. **Accuracy in detecting events**: specifically the type of event. For example, an upward-downward translation should not be reported as a rotation around another axis

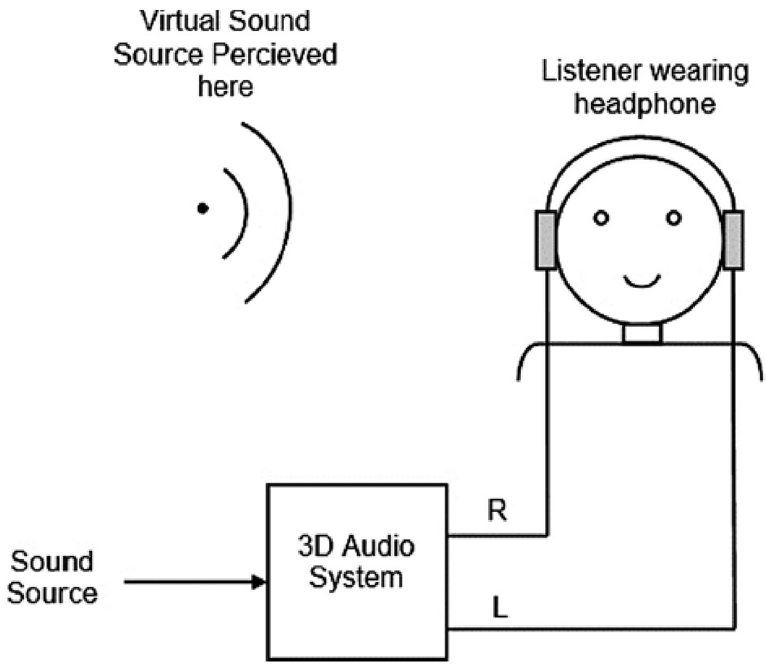



**Figure 1.4:** *Virtual Auditory Display: binaural synthesis. Original picture from [5].*

### 1.2.1 Head Tracking

Visual head tracking systems can be divided in two classes: *marker-free systems* and *marker-based system*. *Marker-free systems OR appearance-based methods* only exploits optical sensors to measure movements of the human head. They acquire the information directly from the recorded images without trying to build a 3-D representation of the users head and they need to acquire also muscle contractions and relaxation (usually more expensive). Viceversa *marker-based systems OR model-based systems* apply identifiers which are placed upon the human head to capture movements. The following list describes the most common methods for capturing this ground truth data, in an approximate order from least accurate (coarse) to most accurate (fine) [18].

- **Directional suggestion**: A set of markers is placed at different locations around a room and each subject is asked to direct his head toward each of the locations while a camera captures discrete images for every location. This technique has two main issues: first, it assumes that each subject's head is in the exact same physical location in 3-D space and second it assumes that a person has the ability to accurately direct their head toward an object and not always is an automatic accurate task.

- **Directional suggestion with laser pointer**: This is similar to the previous one but a laser pointer is affixed to the subject's head for detecting better discrete locations in the room with much higher accuracy from visual feedback. It still assumes that the subjects' heads are located at the same point in space such that direction is equivalent to head pose.

- **Camera arrays**: several cameras at known positions simultaneously capture images of a persons face from different angles. If each subject head is in the same position during capture, this method is pretty accurate.

- **Magnetic sensors**: a magnetic sensor is placed to the head subject and is used to determine position and orientation angles of the head. They are very accurate but some metal presence in the environment can cause wrong data.

- **Ineritial sensors**: inertial sensors utilize accelerometers, gyroscopes, or other motion-sensing devices. It can be coupled with a Kalman filter to reduce noise.

- **Optical motion capture systems**: these technologies are robust, expensive deployments that are most often used for professional cinematic capture of articulated body movement. They are based on reflective or active makers attached to a person followed by infrared cameras and supported by accurate software algorithms. They are the top of the tracking technology.

### 1.2.2 Transfer effects

Tranfer effects are changes of the standard behaviour/skills of a person during a period of time due to an event. Studies [13][20] say that VAD can influence perception skills and can produce changes in abilities using them in real life with a training, creating tranfer effects humans. In this section two examples of how VADs can be a tool for improving or, in alternative cases, worsening skills are presented.

#### 1.2.2.1 VADs For Training Skills

*Akio Honda et al.*[2] [13] did a study in which three different groups of people performed auditory games and were compared to see relsults and skills improvement. The first 10-people group was exposed to listening with their own[3] HRTFs, the second 10-people group used personalized

[2]Some specific concepts in this paragraph are explained in the next sections

[3]Own HRTF means that it is measured on a specific subject creating a new one from scratch. Personalized HRTF means that, using a best match algorithm, will be assigned an already created HRTF set to a selected subject.

HRTFs using the DOMISO (see 1.4.4.1 subsection) technique and the last 20-people group was the control group that localized real sound sources without playing the VAD game. The VAD game was an acoustic version of the *whack-a-mole* game called *Bee Bee Beat*: a honeybee sound appears and the blindfolded listener has to ”hit” while a background masking music is played. The player **tasks** can be summarized as follows: first, the player listens attentively to the sound of the bees. Secondly the player has to detect the sound using his head for better understanding where the sounds comes from. Thirdly the player turns to face a location of the presented target and looks to the sound location and finally can use his hammer to conferme his/her choice. After that, the player receives a immediate vibration feedback signal from the hammer and another honeybee is generated. His/her goal is to ”hit” the honeybee as quickly and correcly as possible.
The two training groups were asked to play the game for seven days (30 min/day) within a 2-week period. The results after these tests reveal that the hits increased and misses decreased with training, regardless of the type of HRTFs: own or personalized (see fig. 1.5).

After a follow-up test, after one month by the training, the results revealed that by playing the VAD game:

- **Hit Rate**: it is increased in the sound localization task for real sound sources by approximately 20%
- **Error Rate**: the vertical and horizontal localization error decreased
- **Tranfer Effects**: they were not diminished, even after 1 month.

This experiment proved that playing the VAD game, sound localization skills had been transferred and that the personalized-HRTFs DOMISO technique is valid and robust due to similar performance between own and personalized HRTFs. In conclusion a VAD game can be an useful tool for training perception and listening skills to use them in real life environment.

#### 1.2.2.2 Ventriloquism

Ventriloquism is a well-known illustration of visuo-auditory integration. It occurs when spatial discrepancies between synchronized auditory and visual events lead to a single unified perception, in which location of auditory events are generally biased by visual events. In a study carried out by *Ludivine Sarlat et al.* [20], listeners were asked to see, in a virtual reality environment, objects with their own sounds but with a disparity of 15 degrees between the visual event and the concurrent auditory stimulation, without listeners knew about this fact. After that

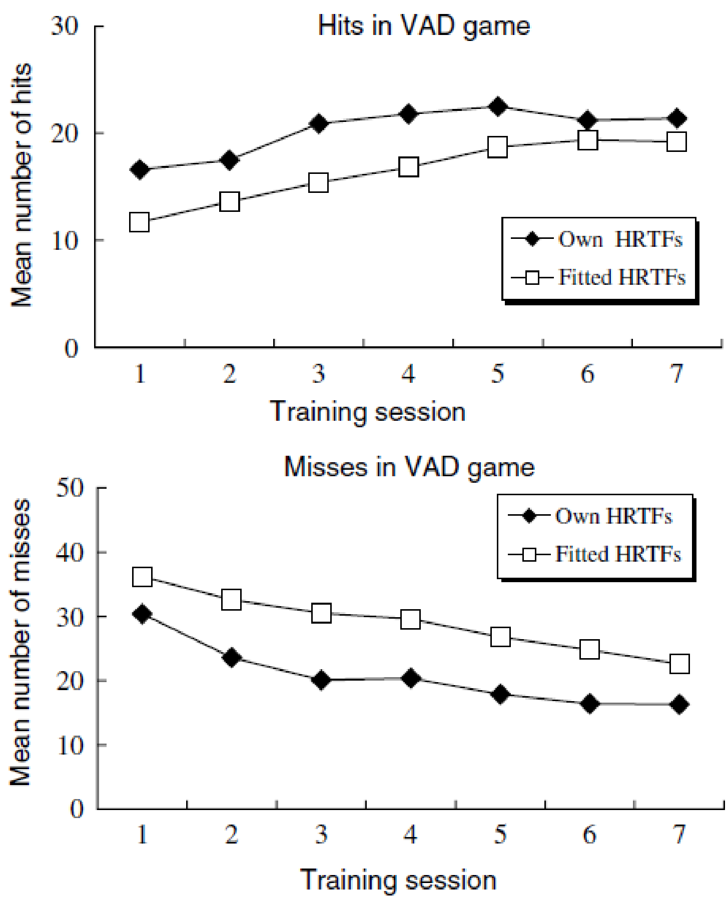



**Figure 1.5:** *Mean hits and misses in the VAD game for own HRTFs and personalized HTRFs groups. The "fitted" word in the plot has the same meaning as "personalized". Original plots in [13].*

"training", the subjects were tested again in not-ventriloquism ambient and *Ludivine Sarlat et al.* noticed that the localization was shifted in both dorsal and frontal hemifields of the listener proving that tranfer effects are occurred and visual stimuli influence the processing of binaural directional cues of sound localization.

## 1.3 General Concepts of Spatial Hearing

In this section, concepts related to *spatail hearing*, a fundamental subject to understand the work showed in this thesis, are presented. Firstly perceptual attributes are explained and some examples are listed of each category. They are followed by the main human body filters and different types of audio systems are decribed. Secondly it is explained how people perceive

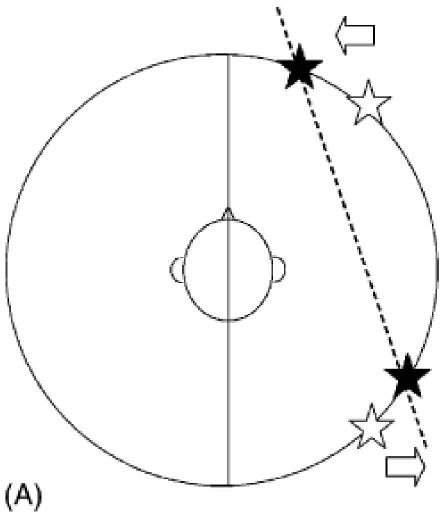



**Figure 1.6:** *Schematic representation of the two different patterns of auditory perception modification: (A) pattern of rotation and (B) pattern of translation. White stars represent perceived locations of auditory stimuli in the pre-training phase and black stars represent perceived positions of the same stimuli post-training. Original plots from [20].*

sound sources and their characteristics [6].

### 1.3.1 Perceptual Attributes

The listener perception of auditory events is related to main groups of perceptual attributes[4]:

1. **Temporal attributes**: rhythm, durability and reverberance are examples of this type
2. **Quality attributes**: loudness, pitch, timbre are examples of this type
3. **Spatial attributes**: direction, distance,spatial impression are examples of this type

The work made and explained in this thesis is about the third group of elements that are related to some effects caused by human body. It is responsible for creating two most relevant cues called Interaural Level Difference (ILD) and Interaural Time Difference (ITD) and monaural cues such as the spectral coloration caused by external ear filtering; these cues can be summarized in the Head-Related Tranfer Functions (HRTFs) for the left and right ear, respectively.

Thanks to individual HRTFs for a specific subject, it is possible to render and simulate personal

[4]Called also elemental senses

filtering for the presentation of virtual sound sources via headphones: this is called *binaural synthesis*. Actually the complete apparatus called *Virtual Auditory Display (VAD)*[5] uses, for each signal, a pair of HRTFs based on the listener-source relation and a pair of *Headphones Tranfer Functions (HpTFs)* based on the acoustic properties of the playback device used by the listener. That is why generally references mention them as ”sets of HRTFs” highlighting the existence of more pairs of HRTFs (left+right) for each available source-listener measurement/simulation.

### 1.3.2 Main Human Body Filters

- **Head**: It is a rigid object between the ears and acts as an obstacle for the propagation of the sound, so it is the cause of two main effects. *Interaural Time Differences (ITD)* is because sound waves have to travel to the other farthest ear and *Interaural Level Difference (ILD)* is due to the ”shadowed” effect by the presence of the head.

- **External Ear**: it is divided in the *pinna* and *ear canal* until the eardrum. After the eardrum there is the middle and internal ear. It behaves like an acoustic antenna capable of amplify some frequencies and attenuate others

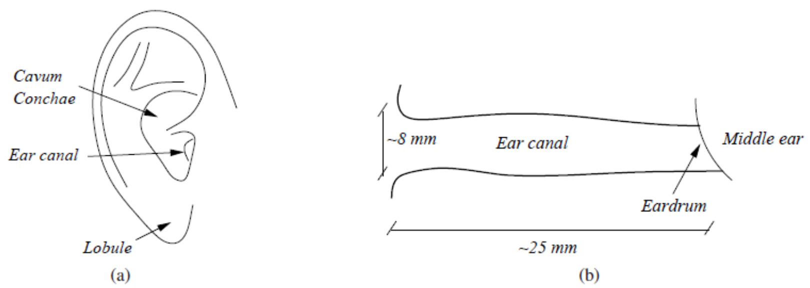



**Figure 1.7:** *External Ear: (a) pinna (b) ear canal*

- **Torso and shoulders**: they contribute to add additional reflections that are added to the direct sound and they have a shadowing effect for sounds coming from below. The most famous way to describe it is the ”snowman model” (see fig. 1.8). Thanks to torso and shoulders, one ear listens to one initial pulse, followed by a series of subsequent pulses that are more delayed when the elevation becomes higher.

[5] See Section 1.2

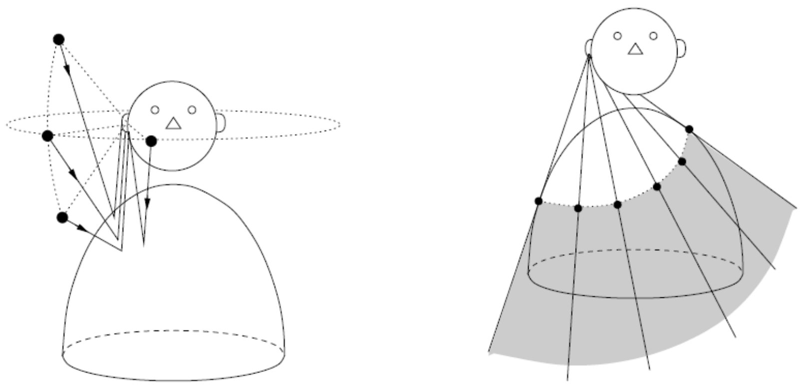



**Figure 1.8:** *Effects of torso and shoulders: (a) reflections (b) shadowing*

### 1.3.3 Perception of sound source location

- **Azimuth**: the **horizontal** perception of sounds has two key components called the ITD and ILD effects, which couple is usually called *Duplex Theory*. People can hardly understand the horizontal direction of a high frequency sine wave, viceversa they can better dectect low frequency sounds (ITD). The IDT at high frequencies is more marked with respect to the low ones. Both effects are exploited together to provide azimuth perception throughout the audible frequency range.

- **Elevation**: there are some conical surfaces called *cones of confusions* where sound sources located on those areas produce identical values of ITD and ILD. In these zones there is an higher front/back confusion because of almost[6] identically values of ITDs and ILDs. Spectral notches can be used to localize sound in elevation through spectral analysis. The attenuation of particular frequencies by the pinna provides information about the **vertical** dimension.

- **Distance**: wihout other information, intensity is the primary distance cue used by listeners who, learning from expeirence can understand how far is the sound source that generates that sound. The perceived magnitude of intensity is called *loudness*. Another aspect that is related with distance is familiarity with the sound: if the subject knows that usually that sound comes from a certain distance, probably when he/she listens to it again will be able to assign a specific distance based on previous experiences.

- **Dynamic**: this is the last aspect which helps listeners to perceive sound sources location:

[6]Due to facial features and asymmetries, they are not usually exactly equals

it is related to the listener position changes.

### 1.3.4 Systems for spatial audio rendering

- **Stereo**: the simplest way to enable spatial sound. Due to the fact that signal is divided in the left and right part, each of them is sent to the appropriate loudspeaker, creating a different perception of the sources of sounds. If listener correctly positioned at the right distance from both of them, he/she can listen to a ”phantom source” in the middle of the two channels, succeeding in listening to all the possible combinations of position sources between the full range Left-Right.

- **Multichannel systems**: this technology is highly advanced and the idea behind that is to have different channels for each possible direction, including above and below. The result can be obtained using several loudspeakers in every needed direction and a bigger one for the non-directional low frequency sounds.

  Both **Stereo** and **Multichannel systems** based on loudspeakers, suffer from the *cross-talk* effect, that is the fact that the sound emitted by one loudspeaker will be always heard by both ears. There are some *cross-talk cancellation* techniques using a pre-process aiming at cancelling the signal that arrives to the other ear.

- **Headphone-based**: usually used in HRTF experiments thanks to their positive aspects, instead of other type of systems, they are different from the previous ones for the nature of how they work. They are small and portable, each ear can have a distinct separate signal helping to design 3-D sound rendering algorithms and they eliminate the room reverberation that could be created by loudspeakers. However, they are unconfortable to wear for long periods and could not have flat-frequency responses maybe compromizing spatialization effects.

## 1.4 How to obtain HRTFs

### 1.4.1 What is a HRTF?

A HRTF is the Fourier Transform of the Head-Related Impulse Response (HRIR) that is the impulse response from the source to the eardrum produced by the sound pressure generated by an arbitrary sound source to the eardrum.

The HRTF is a function of three spatial coordinates and frequency, in fact HRTF filters are unique for every position and angle of incidence. In this context the horizontal and vertical angular coordinates are called *azimuth* and *elevation* noted as $\theta$ and $\phi$, while the radial coordinate is named *range* and noted as $r$. The general writing for a HRTF is as follows:

$$H^{(L,R)} = (r, \theta, \phi, \omega) \tag{1.1}$$

where the superscripts $(L), (R)$ indicate the HRTF at the left or right ear. If $r > 1$ we can say we are in a *far field* and the following notation can be used:

$$H^{(L,R)} = (\theta, \phi, \omega) \tag{1.2}$$

So, formally we can define the HRTF at one ear as the frequency-dependent ratio between the *sound pressure level (SPL)* $\Phi^{(L,R)}(\theta, \phi, \omega)$ at the corresponding eardrum and the free-field SPL at the center of the head $\Phi_f(\omega)$

$$H^{(X)}(\theta, \phi, \omega) = \frac{\Phi^{(X)}(\theta, \phi, \omega)}{\Phi_f(\omega)} \tag{1.3}$$

where $X$ is $L$ or $R$.

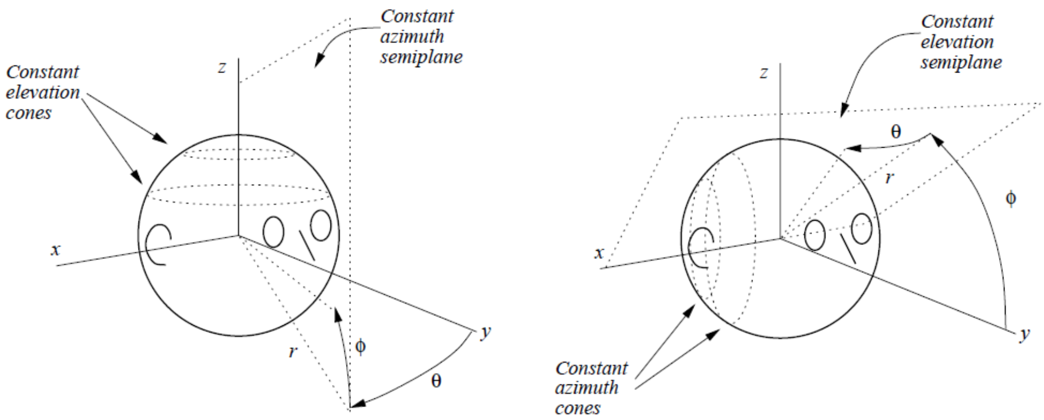



**Figure 1.9:** *(a) Vertical-polar coordinate system (b) interaural-polar coordinate system.*

When HRTFs in a VAD are not personalized to a specific listener, the accuracy of localization is often low and produces large localization errors, typically appearing as frequent front-back confusion and lack of externalization [23].

### 1.4.2 Limitations

One of the main limitations of virtual reality binaural audio personalization and commercialization is the hard work behind the creation of the individual Head-Related Tranfer Functions (HRTFs) that captures all of the physical effects creating a personal perception of audio. The measurement of a listener's own HRTFs in all directions requires a special measuring apparatus and a long measurement time, (see sections below) often a too heavy task to do for each subject involved in an experiment/test. That is why often alternative ways are used to measure HRTFs giving to a subject a personalized, but not individually created, HRTFs set: a trade off between quality and cost of the measure [11] [7].

### 1.4.3 Individual/Own

The standard setting for individual HRTF measurement is in an anechoic chamber with a set of loudspeakers mounted on a geodesic sphere (with a radius of at least one meter in order to avoid near-field effects) at fixed intervals in azimuth and elevation. The listener, seated in the center of the sphere, has microphones in his/her ears. After the set up is complete HRIRs are measured playing analytic signals and recording responses producted at the ears for each position.

The goal is to extract the set of HRTFs for every user who wanted to have their individual/own transfer function. There are some negative aspects in that, such as listener's movements limited to a few positions and the specific locations, that have their own characteristics, used to extract the HRTF sets. Moreover, repeatability of HRTF measurements are still a delicate issue [2].

### 1.4.4 Personalized/Generalized

The personalized HRTFs are chosen among already built HRTFs sets instead of doing the individualized procedure. In fact this procedure is based on the match between tested subjects and database subjects already stored. The most interesting and important part, is the method of how is assigned a particular set of HRTFs to a subject instead of to another set. Researchers are finding different ways to deal with this issue and there are different more or less expensive alternatives using different hardware and/or software tools. Some practical useful methods here are proposed to get personalized HRTFs for tested subjects:

- **DOMISO**[7] Tournament [14] [24]

[7] Determination method of OptimuM Impulse-response by Sound Orientation

**Figure 1.10:** *HRTF measurement setup at the Acoustics Research Institute, Austrian Academy of Sciences, Vienna. Original picture from [8].*

- **Two Steps Selection** [20]
- **Matching anthropometric ear parameters** [25]
- **Drawing Pinna Contours** [12] [19] [22]

#### 1.4.4.1 DOMISO

In this technique [14][24], subjects can choose their most suitable HRTFs from among many, taken from a database following a tounament-style listening tests. The database (corpus) is built using different subjects, storing 120 sets of HRTFs, one set per listener. The method is described as follows:

1. From among all 120 sets of HRTFs, 32 sets of **HRTFs** in the corpus are **selected randomly**.
2. An **orbit** consisting of 13 virtual sound images (one-second-long pink noise, 16 bits, $f_s = 48kHz$) for every 30 degrees counterclockwise from the front of a listener in the horizontal plane is prepared using the selected 32 sets of HRTFs.
3. The outline of this orbit of the virtual sound images is **shown** to listeners as an **illustration** before the listening test

4. **Tournament matches** are scheduled for 32 orbits rendered by the 32 sets of HRTFs.

5. In a session (or a match), the **listener selects one of two orbits** that better resembles the instructed orbit. The selected orbit is the winner. The winner then proceeds to subsequent matches.

6. Finally, **one set of HRTFs wins** the tournament; it is selected as the personalized set of HRTFs.

Actually the tournament is not a pure Swiss-style tournament but it has a small modification. In a Swiss-style tournament, a match is scheduled between same-time winners. The variation of this style avoids the defeat of two strong sets of HRTFs at early stages of the tournament using a "second chance" criterion able to recycle HRTF sets already discarded (see fig. 1.11).

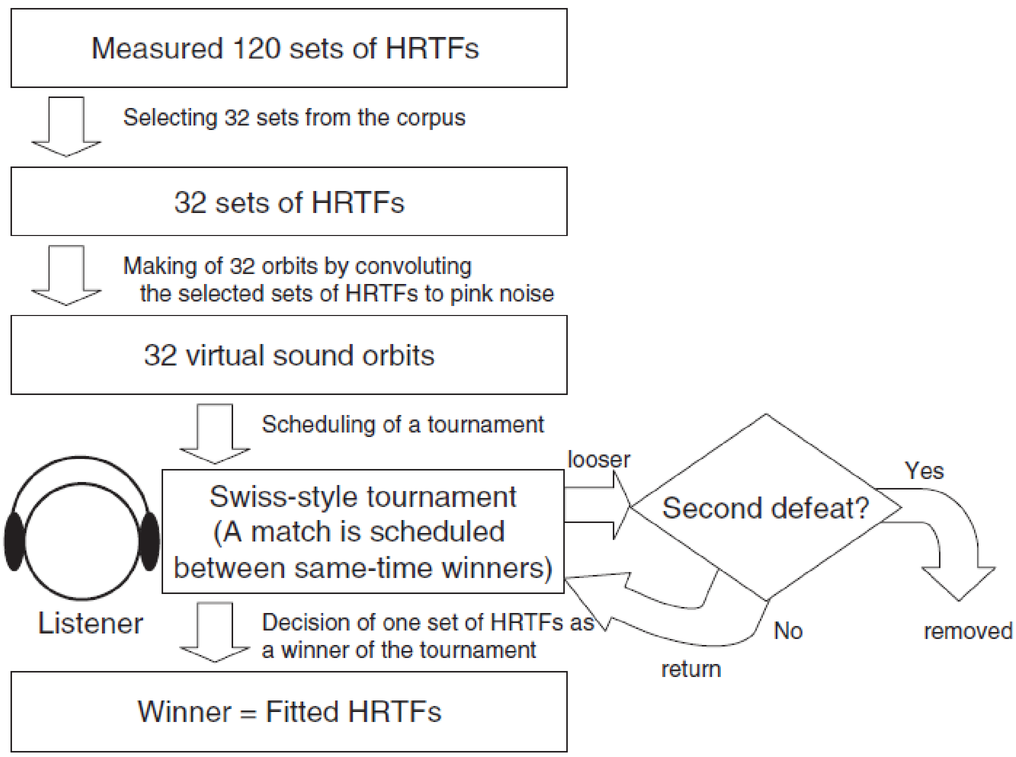



**Figure 1.11:** *DOMISO method for choosing personalized HRTFs. Original picture from [14].*

There are some advantages with respect to measuring individualized HRTFs:

1. **Easy task**: a listener judges the equality of the perceived orbit. Therefore, the task for a listener is simple.

2. **Speedy**: the time cost for individualization is very low (about 15 min), while the measurement time to construct the individualized HRTFs set, required two hours per listener. New techniques were created after this paper for fast individual HRTF measurements that require from $15min$ to $30min$. [17]

The quality performances of this technique were evaluated (see fig 1.12) by *Yukio Iwaya* [14] that proved that the **personalized** DOMISO HRTFs results were similar to **individualized** HRTFs ones but very different from the **away** condition (totally random HRTF, that could not win the tournament).

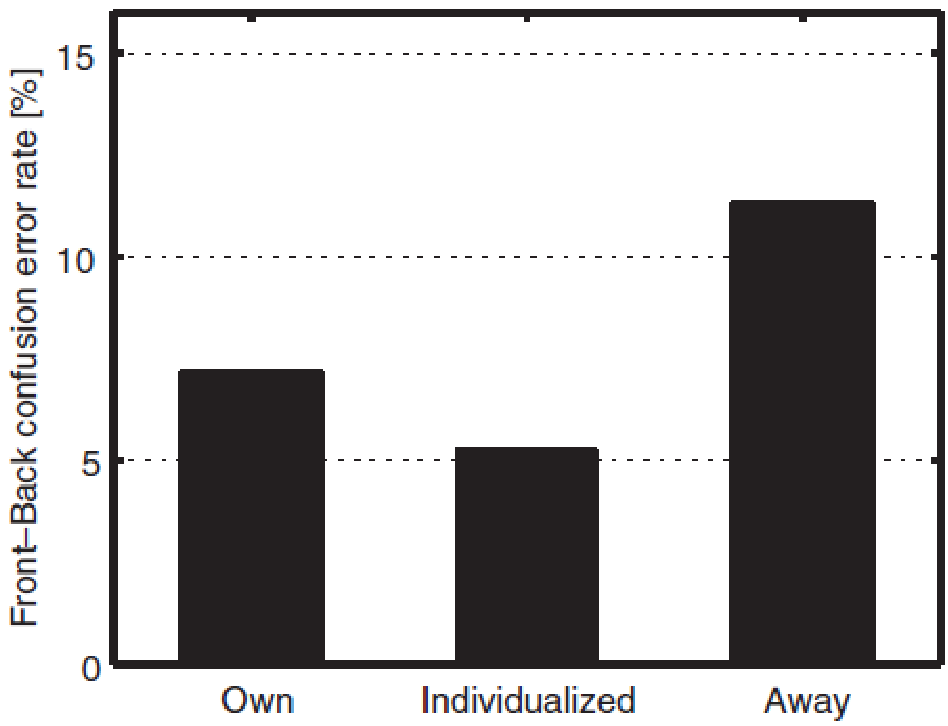



**Figure 1.12:** *Front-Back Error Rate conditions. The word individualized used by Yukio Iwaya in his paper [14] is meant to be "Personalized".*

#### 1.4.4.2 Two Steps Selection

This is a technique based on two different steps. Usually the first step selects from a complete initial pool of HRTF sets, one subset, eliminating the worse ones. The second step selects the best match between the database subject and the tested one operating on the obtained subset.

Different researchers have done several two-steps techniques, each one with its pro and cons. It has been decided to include the Ludivine Sarlat et al. method. used in their work about ventriloquism [20].

The Ludivine Sarlat et al. method is divided in two parts each one done in different days: *top-10 HRTF* selection and *winner*.

In the first part, the subject has to listen to 49 stimuli of white Gaussian noise with the duration of 300 ms, 5ms raise/fall and elevation $0°$. His/her job is to accept or reject stimuli according to the fact that the trajectory has to create a regular and horizontal circle around him/her hearing the sounds outside the head. In this way 10 HRTFs are selected using previous decisions of the subject.

In the second part, for each of the 10 HRTFs previously selected, 16 auditory stimuli similar to the already used ones are presented but convolved at $-45°, -15°, 15°, 45°$ as regards the azimuth and $0°$ as regards the elevation. Like before, the subject has to select HRTFs saying which stimuli are in the front or back hemifield pressing a key.

#### 1.4.4.3 (Matching) Anthropometric ear parameters

This method [25] is based on finding the best match to the external ear shape of a subject using the seven distances, included in the CIPIC[8] database (see fig 1.13). The procedure begins with digital photos of the two ears and the frontal and back side of the body. The seven measurements $d_1, d_2, ..., d_7$ are called in order: cavum concha height, cymba concha height, cavum concha width, fossa height, pinna height, pinna width and intertragal incisure width. After that, the *k-subject* match can be calculated with the following formula:

$$E^k = \sum_{i=1}^{7} \frac{(\tilde{d}_i - d_i^k)^2}{\sigma_i^2} \tag{1.4}$$

where $\tilde{d}_i$ $i = 1, 2, .., 7$ is the value of the $i^{th}$ parameter and $d_i^k$ is the value of the same parameter for the $k^{th}$ subject of the CIPIC database and $\sigma_i^2$ is the variance of the $i^{th}$ parameter across all subjects in the database.

Done all calculations, the best match across all distances is selected as:

$$k = \underset{k}{\operatorname{argmin}}\, E^k \tag{1.5}$$

A Head-and-torso (HAT) model, snowman is the easiest one, can be added to the previous

[8] See next chapter

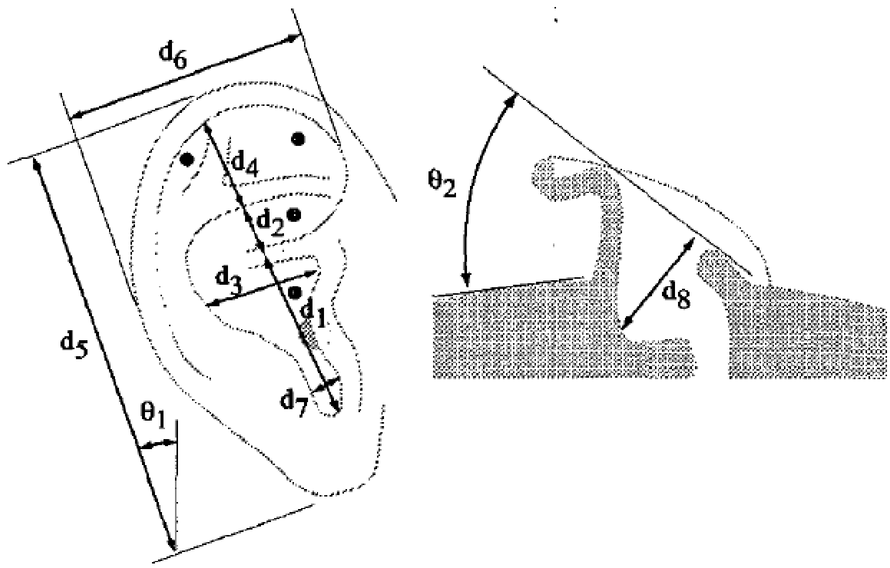



**Figure 1.13:** *CIPIC database pinna measurements. Original picture from [1].*

one to improve results and make it better also in presence of low frequencies.

#### 1.4.4.4 Drawing Pinna Contours

This part is explained in details because it will be the method used to select HRTFs in the chapter 2.The fundamental parts of this method are the pinna anatomy (contours) and acoustic reflections caused by it. In the ray-tracing reflection models, a sound is simplified assuming it like a ray rather than a wave. This simplification is correct as long as the wavelength of the sound is smaller in comparison with dimensions of the involved reflection surfaces, and this is the case where spectral notches due to pinna reflections appear ($f > 6KHz$). In fig. 1.14 a HRIR measurement taken from CIPIC database can be seen where body reflections are visible [19].

The main idea behind the HRTF extraction is that because of reflections and incidence angles in elevation mainly, spectral nothces are generated. Thanks to them it is possible to understand antropometric distances of the pinna and choose the best match between CIPIC subjects and tested one. [12] [22].

#### Reflection Model

The extraction of HRTFs using reflections and contours is based on some physics effects affecting listened sounds. In particular, the distance between the pinna reflection point and the focus

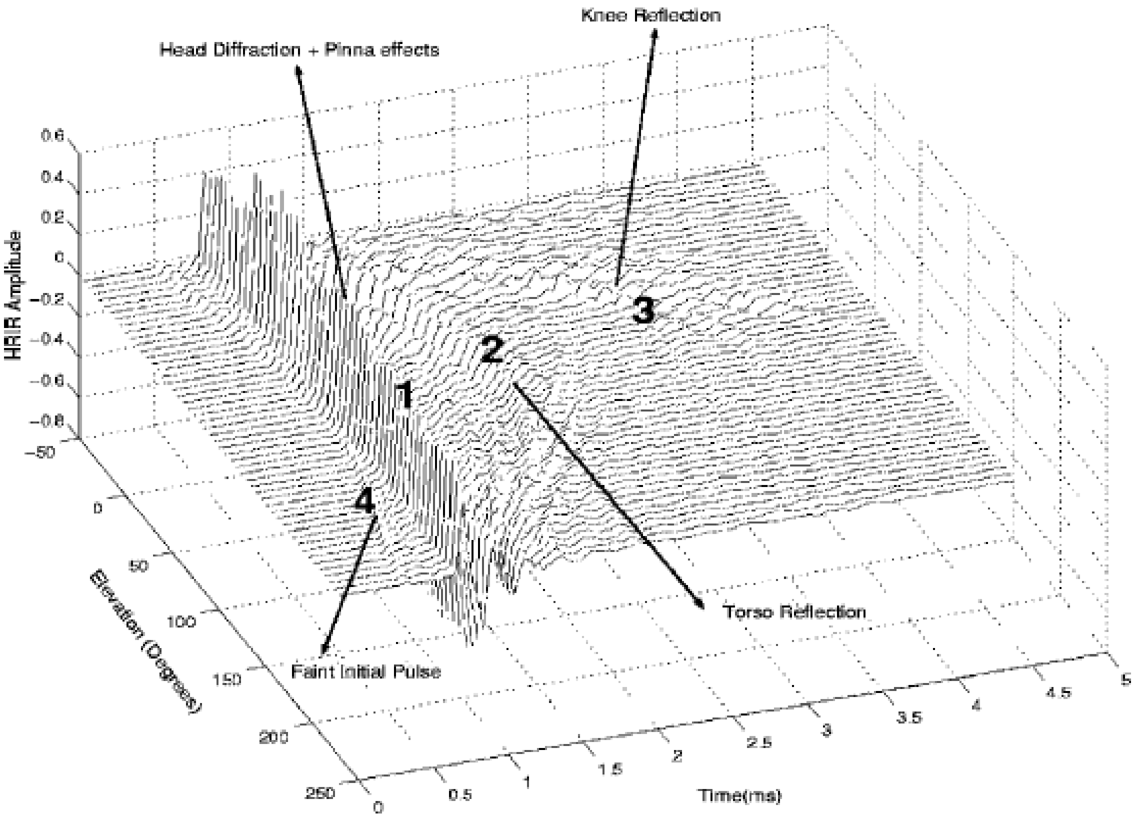



**Figure 1.14:** *The HRIR shown as a mesh plot for the right ear for subject 10 in the CIPIC database for azimuth angle $\theta = 0°$ for all elevations varying from $-45°$ to $+230.625°$. The faint initial pulse, the torso reflection, and the knee reflection can be clearly seen in this plot. Original plot from [19].*

point can be formalized as follows:

$$d_c(\phi) = \frac{ct_d(\phi)}{2} \tag{1.6}$$

where $t_d(\phi)$ is elevation-dependent temporal delay between the direct and the reflected wave and $c$ is the speed of sound.

Assuming the reflection coefficient to be positive, and remembering that the frequency is the reciprocal of the period, a notch is created at all frequencies where reflection's phase shift equal to $\pi$:

$$f_n(\phi) = \frac{2n+1}{2t_d(\phi)} = \frac{c(2n+1)}{4d_c(\phi)} \tag{1.7}$$

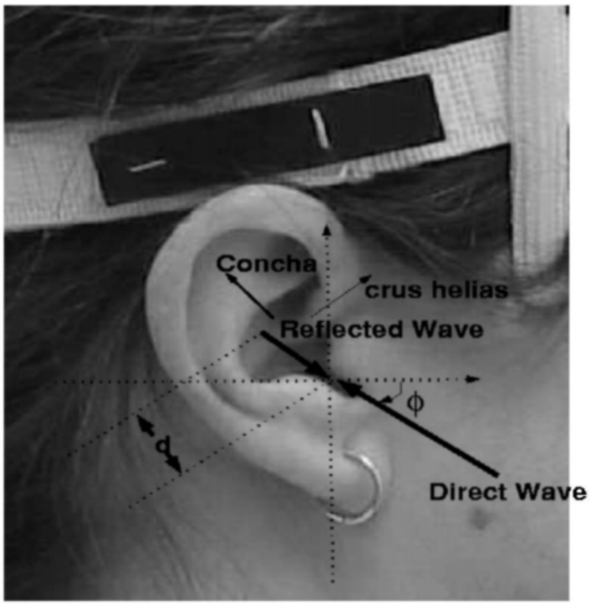



**Figure 1.15:** *Simplified reflection model for pinna spectral nothces. $\phi$ is the incident angle where the direct wave gets reflected from the concha. The delay space is $\mathrm{d}$ doubled because of to and fro of the sound. Original picture from [19].*

where $n \in \mathbb{N}$. Thus, the first notch frequency is found when $n = 0$, giving the following result:

$$f_0(\phi) = \frac{c}{4d_c(\phi)} \tag{1.8}$$

Actually, studies [21] proved that almost $80\%$ of CIPIC subjects exhibit a clear negative reflection in their HRIRs: in fact a more accurate model can be made. When negative reflections appear, notches are found at full-wavelenght delays and not anymore at half-wavelenght ones. That is why, the new reflection equations can be written as:

$$f_n(\phi) = \frac{n+1}{t_d(\phi)} = \frac{c(n+1)}{2d_c(\phi)} \tag{1.9}$$

where $n \in \mathbb{N}$ and

$$f_0(\phi) = \frac{c}{2d_c(\phi)} \tag{1.10}$$

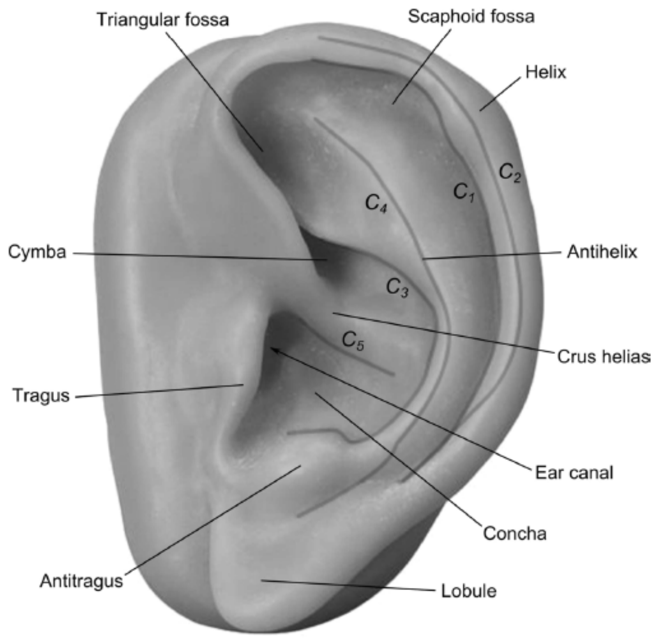



**Figure 1.16:** *Pinna anatomy and the five chosen contours for the matching procedure. Original picture from [22].*

### Pinna Contours

The second crucial element to this extraction method is pinna and its contours. In a human body pinna, five different contours can be identified according to [22] (see fig. 1.16)

1. Helix border $(C_1)$
2. Helix inner wall $(C_2)$, the more external one
3. Concha outer border $(C_3)$
4. Antihelix and concha inner wall $(C_4)$, pretty much in the middle of the pinna
5. Crus helias inferior surface $(C_5)$, the most internal part

After some experiments, the results [22] proved that three main notches are linked to three different reflections on the concha border, antihelix/concha wall and helix.

That is why the previous model of five contours can be simplified and reduced to the three main contours (see fig. 1.17):

- First contour: helix border $(C_1)$

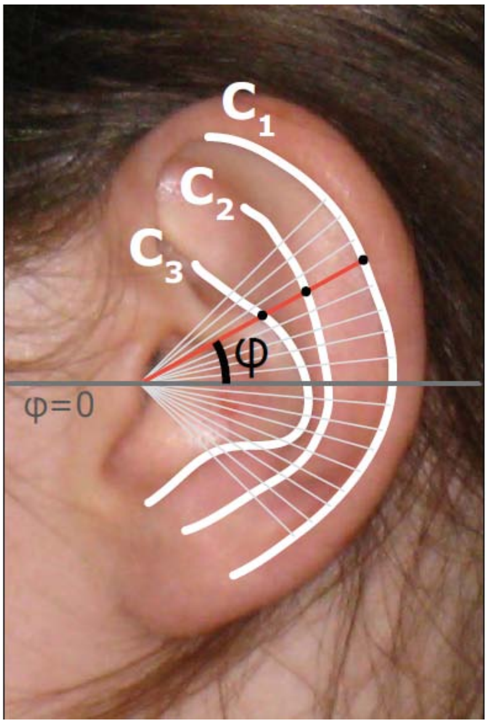



**Figure 1.17:** *Example of the Focus Point and the three main Pinna Contours. Original picture from [12].*

- Second contour: antihelix and Concha inner wall $(C_2)$
- Third contour: concha outer border $(C_3)$

### Subject Mismatch

Based on the previous concepts, the HRTF mismatch between two subjects can be defined by the following equation:

$$m = \frac{1}{n} \sum_{i=1}^{n} \frac{w_i}{N_\varphi} \sum_{\varphi} \frac{|f_0^i(\varphi) - F_0^i(\varphi)|}{F_0^i(\varphi)} \tag{1.11}$$

where $n$ is the number of contours that are considered, $w_i$ is a convex combination of weights and $\varphi$ spans all the $[-45°, +45°]$ frontal elevation angles for which the $i^{th}$ notch is present in the corresponding HRTF and $N_\varphi \in \{1, 2, ..., 17\}$ is the number of actual sums done in the internal $\sum$, i.e. when $f_0^i(\varphi) \neq 0 \wedge F_0^i(\varphi) \neq 0$. $\varphi$ assumes 17 different values on steps of $\frac{90°}{17-1} = 5.625°$ each: it is worthwhile to mention that this is because those 17 steps are the most informative ones as regards frequency notches. $f_0^i(\phi) = \frac{c}{2d_c(\phi)}$ are the frequencies extracted from the image and contours of the tested subjects and $F_0^i$ are the notch frequencies extracted of the available median-plane HRTFs thanks to the *structural decomposition algorithm* [10].

A subject will receive the HRTFs set with the lowest $m$ among all available HRTF notch tracks.

In the next chapter it is shown how to adapt these concepts and apply them to a matlab application that will be able to find the best match between a CIPIC ID and a selected subject.

# Chapter 2

# Extraction of HRTFs using Matlab

The first task done was to modify and enhance an already existing matlab program created to select the best HRTFs set from the CIPIC database, starting from pinna images of the selected subject.

The goal of this task is to have a better HRTF set to be loaded in the virtual reality environment and obtain better results, especially in vertical dimension, rather than using, for example, the generic KEMAR functions (CIPIC ID 21,165).

The real need of this approach was to give to unskilled people too, the possibility to extract and use, in a efficient way, personalized HRTFs. That is why a complete, light and easy-to-use graphic program, provided in a package/installer, has been created.

#### CIPIC HRTF Database

One possibility for using public databases is the CIPIC, created at the U. C. Davis CIPIC Interface Laboratory. The first release had HRIRs for 45 subjects at 25 different azimuths and 50 different elevations, for a total of 1250 directions (see fig 2.1). In addition, the CIPIC database includes a set of anthropometric measurements useful for scaling studies (see fig 1.13) [1].

## 2.1 How To Use It

This section describes only the features of the program from the point of view of the user. All technical details are proposed to the next sections.

A complete installation package of the application is provided to the user and it needs to be installed using the Matlab Runtime. After the installation, the program is ready to be used.

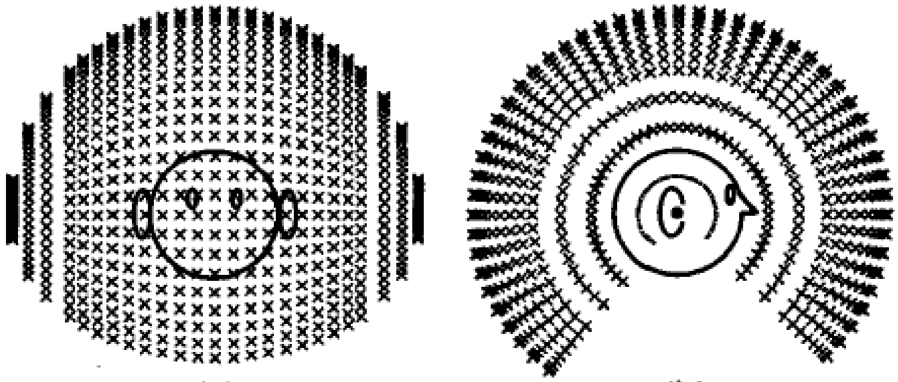



**Figure 2.1:** *All locations of data points in the (a) front and (b) side of CIPIC subjects. Original picture from [1].*

### Subject Information

The main GUI called ”Subject HRTF selection” has the task to manage subjects, all of them organized using a list on the left of the screen (fig. 2.2). Thanks to the three different buttons ”Add Subject”, ”Update Subject” and ”Delete Subject” and some text-fields used to assign to each subject his/her own information, it is possible to manage the subject list in an efficient and organized way.

For each subject stored in the list, an image of the left ear can be assigned with the button ”Choose ear image”: the image will be shown in the middle of the GUI when a name from the list is clicked. In case of a subject without a loaded ear image, a ”No Image” writing appears instead of it.

### Trace Ear

After loading the personal left or right ear image of a subject, it can be traced clicking on the button ”Trace Ear”. Done that, two parameters are loaded: the ”Contours” and ”Ear Canal” ($N$ and $K$) text-fields are passed to the new GUI that is opened on the click of the ”Trace Ear” button (fig. 2.3). In the trace-ear GUI the user has to draw by hand $N$ $C_1$ contours following the image as guideline. After that, the user has to point $K$ positions of the ear canal entrance. The main reason of drawing and pointing several times the same object is twofolded: inaccuracy by the user, and the inherent ambiguities in the 2D picture (contour not visible, ear canal entrance not well defined). Thus, in this application, it is possible to use average positions instead of trying to find the optimal one. A pixel-to-meter tool is used to convert each cropped image

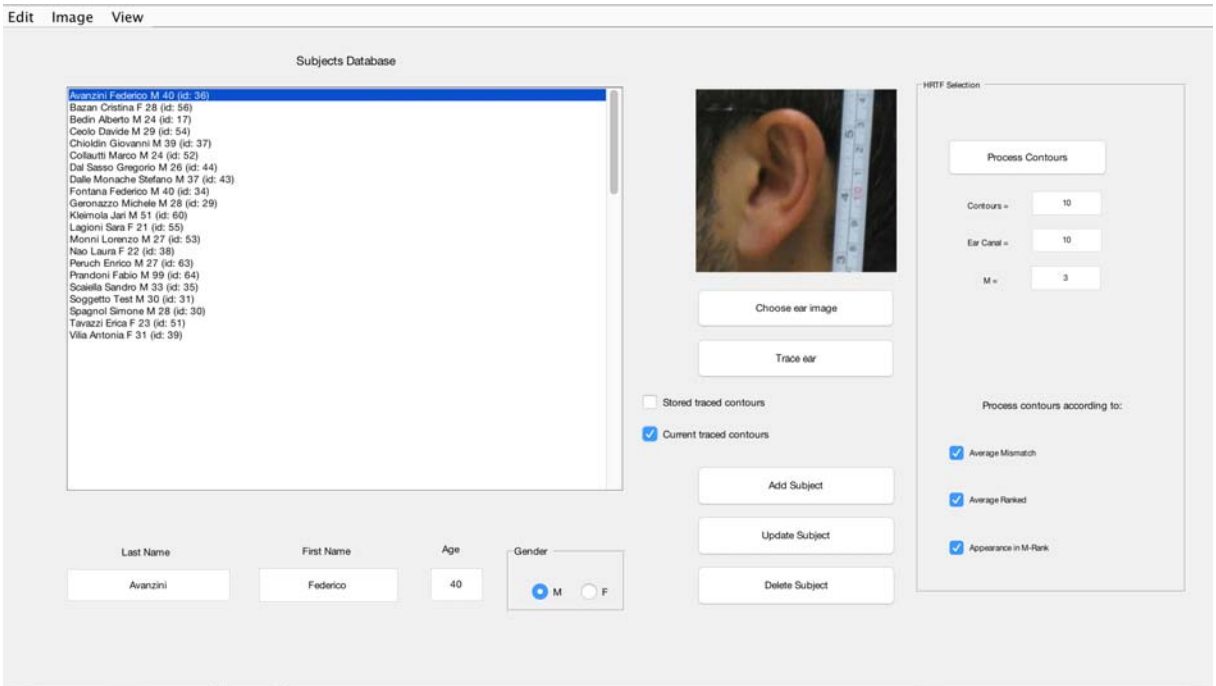



**Figure 2.2:** *Main Matlab Application GUI*

lenght and make them possible to be compared.

In this GUI there are two checkboxes to improve the usability of the tracing task: the first one is the "Stored Traced Contours" that shows the already drawn contours in the previous draw session. The second one called "Current traced contours" is about not erasing the current contours already drwan in the current session.

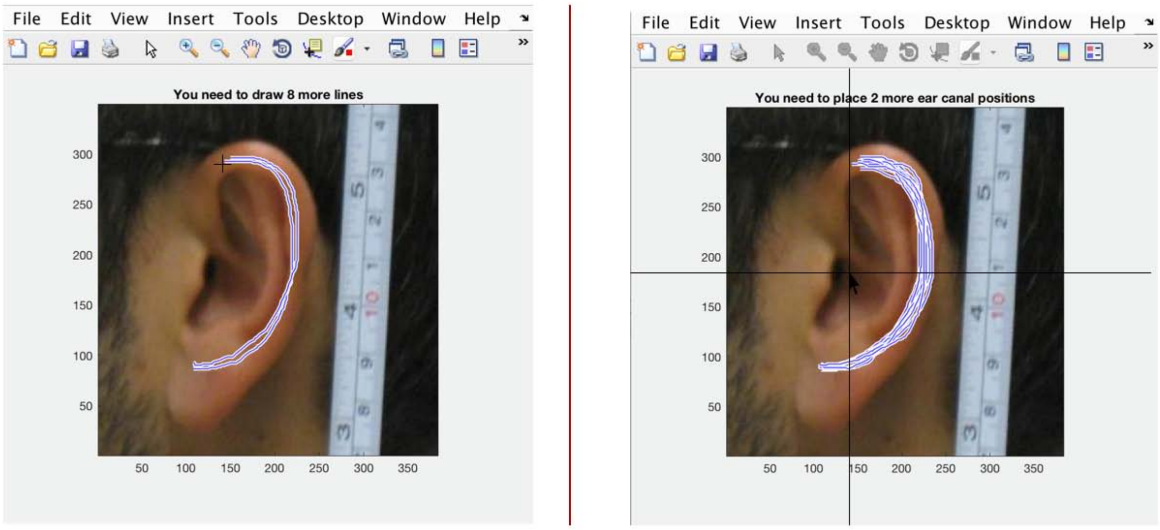
### Results

The last step necessary to obtain the best match between the tested subject and the CIPIC is to click "Process Contours". The matlab application will process all the subject's data and give as output three *.pdf* files, each one related to the three classifications seen before[1]. Moreover a fourth score *.pdf* file is processed and given to the user as the best match classification. The lowest score assigned to the CIPIC ID will be the best choice.

### Additional Features

Clicking on the toolbar of the main GUI, it is possible to view all the four *.pdf* result files of a previous tested and selected subject in the list and to change the default parameters like the

[1] Average Mismatch, Average Rank, Appearance in Top-M.




*(a)* *(b)*

**Figure 2.3:** *Trace: N-Contours (a), K-Ear Canals (b)*

pixel-to-meters factor, $K$ and $N$. Moreover, some additional useful operations like *clear the database* and *rotate images* can be performed.

## 2.2 Technical Details

This section explains in datail how the Matlab code extracts the information needed to assign to a tested subject a CIPIC ID. This method is described in section 1.4.4.4 as a quality improvement and easy of use. Michele Geronazzo et. al. [12] studied the three main contours and their associated reflections, concluding that the best results criterion is to assign the whole weight to $w_1 = 1$, i.e. considering only the $C_1$ contour. Thus, the formula of the mismatch becomes:

$$m_{(k,n)} = \frac{1}{N_\varphi} \sum_{\varphi} \frac{|f_0^{(k,n)}(\varphi) - F_0(\varphi)|}{F_0(\varphi)} \tag{2.1}$$

where $(k, n)$ with $(0 \leq k < K)$ and $(0 \leq n < N)$ is referring to a one particular couple of traced contour and ear canal position. To create a complete reliable HRTFs extraction tool, the user has to draw $N$ times the $C_1$ contour and $K$ times the focus point/ear canal. This multiple tracking improvement increases the robustness of the application and decreases drawing human faults. In fact an average and standard deviation calculations are needed to give reliable and complete results, processing all possible mismatches linked to $K \times N$ pairs $(k, n)$. When the user has drawn $K \times N$ objects, $\boldsymbol{f_0}$ is calculated and stored into a file. It is a $K \cdot N \times 17$ matrix where the rows are the experiments and the columns are the number of elevations of $\varphi$ utilized in calculations. In the 'Data Processing' phase, it will be compared to $\boldsymbol{F_0}$, another matrix, extracted from each CIPIC subject HRIRs, that contains all the available notch tracks.

### 2.2.1 Data Processing

To give a solid ranked scored list of CIPIC IDs as output, it has been decided to use three different less and less precise ordered classifications and calculate the score as a linear combination of these. Each one is created by one of the three basic matrices below (each represented by a[2] $K \cdot N \times 43$ matrix):

1. **Mismatch Matrix**: the most specific one gives, for each CIPIC ID, the mismatch assigned to each pair $(k, n)$, calculated with the equation 2.1.
2. **Ranked Matrix**: the intermediate of the three classifications, it gives the ranked position of each CIPIC ID, for each pair $(k, n)$. It is created from the *Mismatch Matrix* by sorting each row as an indipendent array and substituting each mismatch value with the ranking position of the sorted array.
3. **Top-M Matrix**: the less informative and more general matrix is based on the $M$ parameter, that counts, for each pair $(k, n)$, how many times that CIPIC ID is in the top-M ranked classification. It is created from the *Ranked Matrix* by substituting each ranked position with a $1$ if it is in the Top-M classification, $0$ otherwise.

Having obtained all the three basic matrices, some operations can be done on those structures to create a better compact view of the results. For each basic matrix, a new ordered overall version is calculated, allowing to understand the global results about a CIPIC ID. In particular the three new matrices are the following:

1. **Average Mismatch**: created from *Mismatch Matrix* doing calculations by columns, it has $43$ rows like the number of CIPIC IDs and $3$ columns that stand for *Average Mismatch, Standard Deviation* and *CIPIC ID*
2. **Average Rank**: created from *Average Matrix* doing calculations by columns, it has $43$ rows like the number of CIPIC IDs and $3$ columns that stand for *Average Rank, Standard Deviation* and *CIPIC ID*

[2] Number of CIPIC subjects with the $C_1$ contour notches information

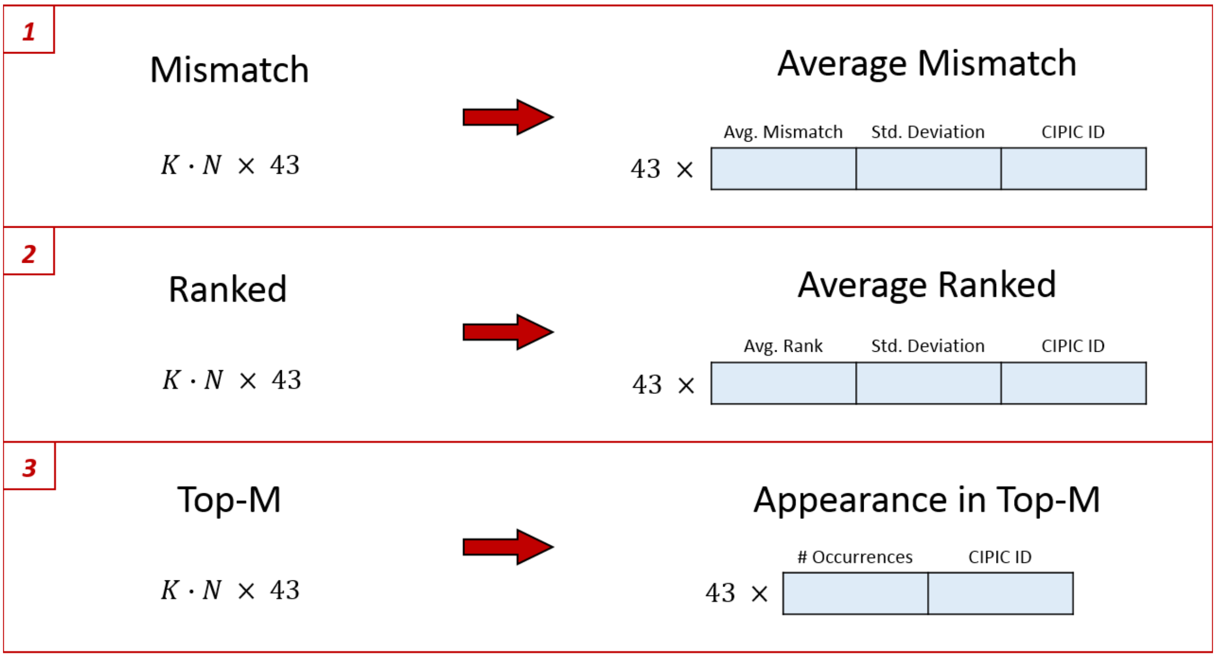



**Figure 2.4:** *Processing Matrices*

3. **Occurences in Top-M**: created from *Top-M Matrix* doing calculations by columns, it has $43$ rows like the number of CIPIC IDs and $2$ columns that stand for *Number of occurences (in Top-M)* and *CIPIC ID*

### 2.2.2 Scoring CIPIC IDs

The final score assigned to a CIPIC ID is given by a combination of each score of the three above matrices (see below).

It has been decided that each matrix score is assigned in the following way:

1. **Average Mismatch**: the CIPIC ID at the $i^{th}$ position gets $i$ points

2. **Average Rank**: the CIPIC ID at the $i^{th}$ position gets $i$ points

3. **Appearance in Top-M**: the scoring of this classification can be seen in fig. 2.5.

It's important to highlight that this is a crucial point of this work because there can be lots of different possible weighted ways of assigning scores able to modify the final results and HRTFs choices. This method has been chosen due to its balanced and rational nature of scoring each matrix and the way of finding the final score (see below).

$$
\begin{array}{l}
points \leftarrow 1 \\
for\ (i \leftarrow 1\ to\ 43) \\
\quad CIPIC_i \leftarrow points \\
\quad if(NumOcc_{i+1} > NumOcc_i) \\
\quad\quad points \leftarrow points{+}1
\end{array}
$$

**Figure 2.5:** *Scoring Occurrences in Top-M*

In the main GUI there are three different checkboxes to select which classifications the user wants to include in the score. Let be $B_Y$ a 'boolean' value that is $1$ if the checkbox $Y = 1, 2, 3$ is selected and $0$ otherwise. The final score is assigned to each CIPIC ID following this criterion:

$$S = \sum_{Y=1}^{3} S_Y \cdot B_Y \tag{2.2}$$

where $S_Y$ is the score of a CIPIC ID given by the $Y$ *matrix*. The best CIPIC ID is the one with the lowest score.

# Chapter 3

# Personalized-Audio VR Tools

In this chapter different tools used in this project for the binaural audio rendering and environment creation are explained. The first part is about the installation and utilization of HOBA. Technical details and the choices on the market for the VR implementation then follow. The block scheme of the applications is depicted in fig. 3.1.


**Figure 3.1:** *Block scheme of application tools.*

## 3.1 HOBA: HRTFs On-demand for Binaural Audio

HOBA (HRTFs On-demand for Binaural Audio) framework enables custom HRTFs in browser spatial audio rendering applications. Not only some free software applications are required to be downloaded, but the HOBA web application for the audio rendering is also needed. Prerequisites will be examinated and illustrated in the following subsection.

### 3.1.1 Prerequisites

Before running the VAD system it is needed to install and configure a bridge that will be used to send listener's data to a web browser application and to render the HRTFs based on those information.

#### Node.js and nmp

Node.js is a light and efficient runtime JavaScript environment, built on Chrome's JavaScript V8 engine. It's open source and usually used to develop a variety of server tools and applications. The OSC bridge is based on that environment so it is required to be installed.

#### Open Sound Control (OSC)

Open Sound Control is a protocol for muscial performance or show control that is a useful tool for interoperability, accuracy and flexibilty. In order to install the bridge, it is required to use *nmp* package manager for JavaScript using the following commands given by the shell or Windows prompt:

- cd osc2ws
- npm install ws
- npm install osc

Once it is accomplished, in order to run the HOBA application, the starting command *node osc2ws.js* is needed.

By default, OSC UDP messages received from port 7400 are routed to the browser websocket port 8081. These parameters can be changed in the *osc2ws.js* source code.

After setting up the bridge, the web application *hoba-oscdemo.html* must be run by a browser (Mozilla FireFox suggested, see fig. 3.3). After that, the message *A Web Socket connection has*

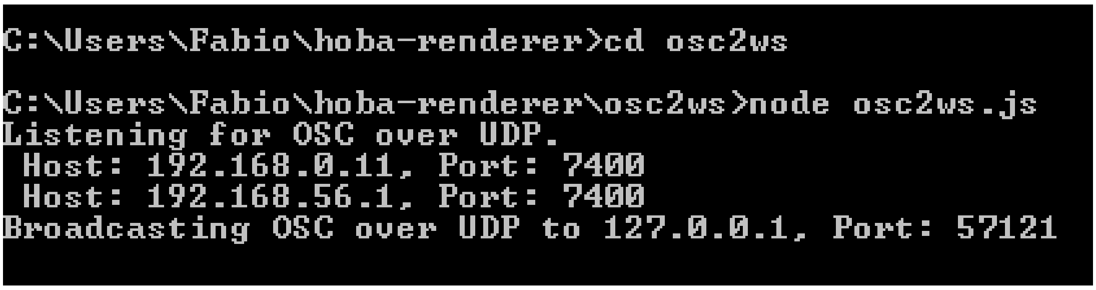



**Figure 3.2:** *Launch The Bridge*

*been established!* is shown in the previous shell/prompt screen and everything is ready to begin the customized audio render!

### 3.1.2 HOBA Application

The HOBA web application (see fig. 3.3) is divided in three parts. The left part is called *soundpool* and it allows to choose a sound source to be rendered, from a pool of *.wav* files. The central part is called *sources* and it loads selected sources and make them play using seven cursors. The right part is called *listener* and thanks to nine cursors it manages the position of the listener-sources.

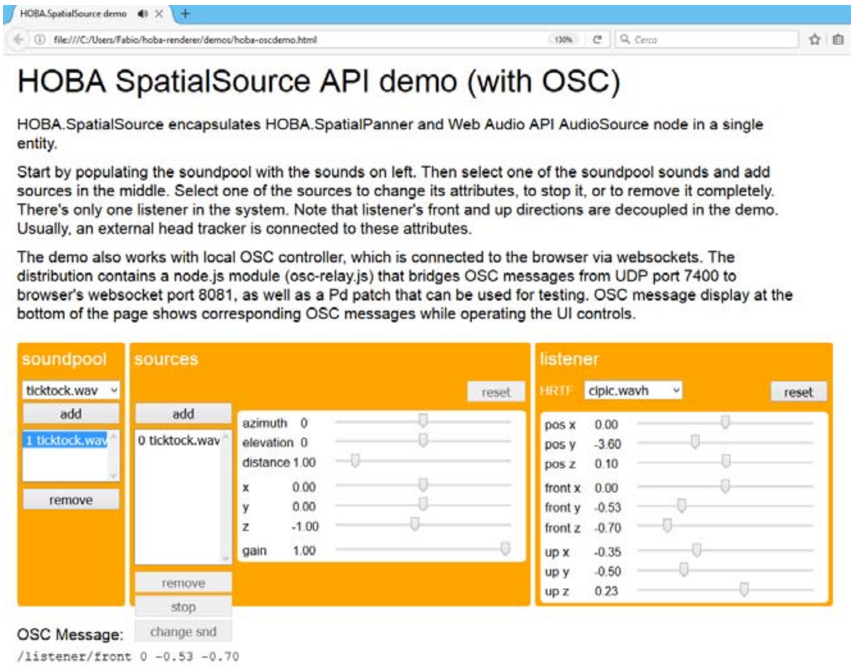



**Figure 3.3:** *HOBA Demo Running*

### 3.1.3 From Sofa to WAVH

Spatially Oriented Format for Acoustics (SOFA) is a file format for storing spatially oriented acoustic data. It is flexible to describe multiple conditions (listeners, distances, etc..) in a single file, supporting compression for efficient data transfer and storage [8]. It aims at storing data representing HRTFs in a general way powering more complex data like binaural or directional room impulse responses (BRIRs, DRIRs). All the CIPIC subjects SOFA files are downlodable from the official SOFA site[1].

#### Conventions

SOFA supports different types of conventions, that are basically specifications of how data and metadata have to be stored. One of the most important conventions, useful to this work, is the *SimpleFreeFieldHRIR*. It considers measured HRIRs with an omnidirectional source for a single listener. This convention describes source position, such as azimuth, elevation and radius, and other important information e.g. variation in head tilt or listener position with respect to the center of the measurement setup.

#### WAVH

In order to have personalized HRTFs, the *CIPIC-Subject-XYZ-Hrir-final.sofa* files are processed one single time each to be converted in the WAVH file format: they will be used in the web browser application to render sounds.

One option to encapsulate this data into WAVH files is to use a chunk-based *Resource Interchange File Format* (RIFF). The WAVH structure includes chunks inside it and each of them has the fundamentals stored data: *elevation, azimuth, distance, itdL, itdR*.

Two advantages are brought by using WAVH format: WAVH files are already supported in WEB Audio API and the chuncked nature of his implementation makes it easier to expand for future updates and specs [8].

### 3.1.4 HpTFs and Headphones

Among all possible choices, it has been decided to use the Sennheiser HD600 headphones thanks to their open nature and flat frequency response.

In order to balance the sounds produced by the VR application and rendered by HOBA, it was decided add an equalizer to the rendering system. Equalizer APO is a parametric/graphic

[1]https://www.sofaconventions.org/data/database/cipic/

equalizer for Windows able to add high and low pass filters in addition to a specific Headphones Transfer Funcion. The specific HpTF (see fig. 3.4) was added to the system as first step. After that, the high-pass filter was set to $159,97Hz$ and the low-pass filter was set to $20KHz$. Finally, a preamplification of $17.80dB$ was added to make the sound clearer and louder for the subjects.

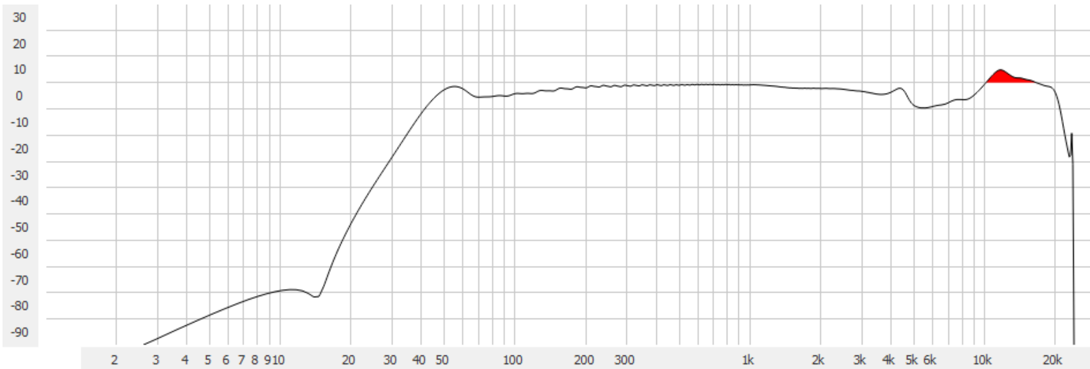



**Figure 3.4:** *Sennheiser HD600 HpTF loaded in the APO Equalizer.*

## 3.2 VR in Unity

Before starting to explain how we created our framework for the experiment, it will be shortly described how each piece of the puzzle is used to create the ending scene. Starting by giving the idea behind a game engine and continue with a comparison of popular game engines, Virtual Reality will be explained and analyzed in its shapes.

### 3.2.1 Software: Game Engines

The videogames industry has grown exponentally in the last years and several of firms have decided to invest in that direction for their future. Because of the internet and an average ”fast” connection, people have the possibility to learn at low costs more than ever before, thanks to official or unofficial good quality tutorials and courses. That is why, some game engine companies have decided to make their full framework available to people in different subscription forms *(Personal, Plus/Base, Premium/Pro)*.

These are names to identify which subscription suits better (price, service) for those who want to begin to create their own game. The *Personal* licence is suggested to people who are not interested in selling their applications in markets or in any other commercial form. The *Personal* licence has limitations in how applications can be developed. The *Plus/Base* licence is for people that want to earn money from their games or are using the engine in their own company for business: it has some advantages compared to the first one but still there are some limitations compared to the *Premium/Pro* licence. This last one, in fact, offers what a person or a company needs in order to begin programming their own complete game with special discounts, unlimited earning possibilities, access to private source codes, special supports, etc...

#### What is a Game Engine?

A game engine is basically a complete software framework that permits people that are using it to create games in a easy way, focusing more on the fun and the gameplay part instead of low level details. Besides, game engines exploit programming code scripts that can already be written in some downloadable *assets* following easy tutorials or forums.

#### Comparison of Engines

Here is the list of the most popular game engines:

- **CryEngine**: created by Crytek GmbH, it is a supreme visual game engine powering the famous *Far Cry* and *Crysis* series of games
- **Frostbite**: from EA Dice, used to power the *Battlefield* and the new *Battlefront star wars* series
- **Gamebryo**: used for *Fallout 3* & *Fallout: New Vegas*
- **IdTech**: created by idSoftware, one of the oldest engines, appeared in *Doom* series in 1993
- **IWEngine**: based on IdTech, used mostly for *Call Of Duty* game series
- **Unity**: popular engine for a high-quality indie (indipendent, not supported by publishers) and mobile games
- **Unreal Engine**: created by Epic Games, used in the *Unreal Tournament*, *Mass Effect* and *Gears of War* series

As already said, few of them, in particular *Unity*, *Cry* and *Unreal* are the famous ones opened to personal developement (paid or free) just with an online subscription. So, how can we choose among them? In section 3.2.3 pro/cons and engines choice will be explained.

### 3.2.2 Hardware: How to run VR

#### VR in the Market

There are lots of types of VR consoles and headsets. The top level visors used by console/PC and the most famous and prestigious are Oculus Rift, PlayStation VR and the HTC Vive.

- **Oculus Rift** was developed by Palmer Luckey and funded via Kickstarter. The Rift plugs into your computer's DVI and USB ports and tracks your head movements to provide 3D imagery on its stereo screens.
- **Playstation VR** is of course created by Sony Playstation 4 and it is a good compromise for quality and price, giving a complete affordable not-top-level visor to everyone who can not afford an expensive one.
- The **HTC Vive** is made in collaboration with PC games giant Valve. It works with Valve's Mammoth gaming ecosystem and it achieves high quality at a high price.

A second category of headsets is represented by those compatible with smartphones. The most famous headsets are: Samsung Gear (2016), Google Cardboard and Google Daydream view.

- **Samsung Gear** is a low cost visor (but expensive compared to the other visors of this category) compatible only with the top level smartphones of that company: Galaxy Note 7, S7, S7 Edge, Note 5, S6, S6 Edge or S6 Edge+. It has also a touch controller to interact with VR applications.
- **Google CardBoard** is the cheapest visor. It is made by unexpensive components and it could be the best choice to join the Virtual Reality world.
- **Google Daydream** view is one of the best quality/price models.

### 3.2.3 Selected Components

#### Software choice: Unity

The choice was among the above three engines. Which is the best for this thesis purpose? The choice must meet these requirements and characteristics:

- **Easy** to use and learn

- **Free** and without too big limitations
- **VR** compatible with the cheapest headsets (VR visors).

*Cry Engine* has only paid subscriptions at the moment, so it was not the best choice. *Unreal* is pretty famous for heavy console and PC games compared to *Unity* which is famous online for supporting instead indie games. Actually both have some little pros and cons, but nothing really relevant. However at the moment *Unity* appears to be the most widely used engine in academic research. In particular, there was the opportunity to access some *Unity* applications already tested by collegues from Alborg University Copenhagen (AAC): it was easier to learn from those than creating new ones from scratch.

#### Hardware choice: Android Smartphone Galaxy S7 and Gear VR

The hardware was limited by an average budget and the idea of finding a lot of possible applications using the same device. Consoles and PCs were too expensive and too specific for the purpose. A great device combination in the smartphone category is the Samsung Galaxy S6-S7 and the Samsung Gear VR headset able to render great android applications and make the user feel very confortable with virtual environments. That is why that specific combination was chosen to be used in the experiment, having the possibility to buy a Samsung Galaxy S7 instead of a Galaxy S6.

# Chapter 4

# The Experiment & Results

## 4.1 Localization Test

A localization test is a task that implies localize sounds coming from different directions. We evaluated localization performances of 15 subjects. The idea behind the experiment was to test three different HRTF sets (A,B,C) and compare them in order to give a ranked classification for the HRTF selection process.

All the tested subjects were in a silent booth in three different days of the same week for an average of $50min$ per subject. Each subject had to complete three different sessions using, without knowing the order, the A, B and C HRTF conditions.

### 4.1.1 Apparatus

#### HW/SW Setup

The experimental setup is based on all HW/SW tools presented in chapter 3. The Samsung Galaxy S7 is coupled with the Samsung Gear VR for the 3D video rendering. A PC is running the complete HOBA environment receiving using a wireless communication through a router the coordinates pointed by the tested subject in the Unity application. All the sounds are listened using Sennheiser HD600 headphones.

#### Unity Scene

Inside a $1m$ sphere, the subject was free to look around and select using a virtual laser pointer, moved by the head, the supposed directions of the generated sound sources tapping on the Samsung Gear VR touchpad. The surrounding environment included a blue sky and a green

**Figure 4.1:** *Experimental Conditions: the subject was closed in the silent booth with the light turned off for the entire experiment*

grass garden. After each tap, the user had to point again on the origin of the sphere, a brown cube at the position $(0, 0)$, waiting for two seconds of silence and follow again the burst sound for another three seconds, iterating the procedure.

The localization test procedure is based on the guidelines provided in a previous paper [12].

Each subject has to complete a block of stimuli in order to proceed with the following one, changing HRTF set, doing a $3$ minutes of pause between each block to decrease the possible fatigue of the experiment and fill the session questionnaire.

After $3$ blocks with three different HRTF sets, the results are ready to be processed by some scripts or excel files. Session data are stored in the smartphone storage using the *.txt* format: the filename is the number of the experimental subject followed by the condition used.

### 4.1.2 Stimuli

#### Sound Sources

Each of the three experimental blocks were made on several trials. In particular a block is a series of $25$ stimuli composed by a train of three $40$-$ms$ gaussian noise bursts with $30$-$ms$ of silence between each burst, repeated six times (to a total of $3s$). This sounds have been proved

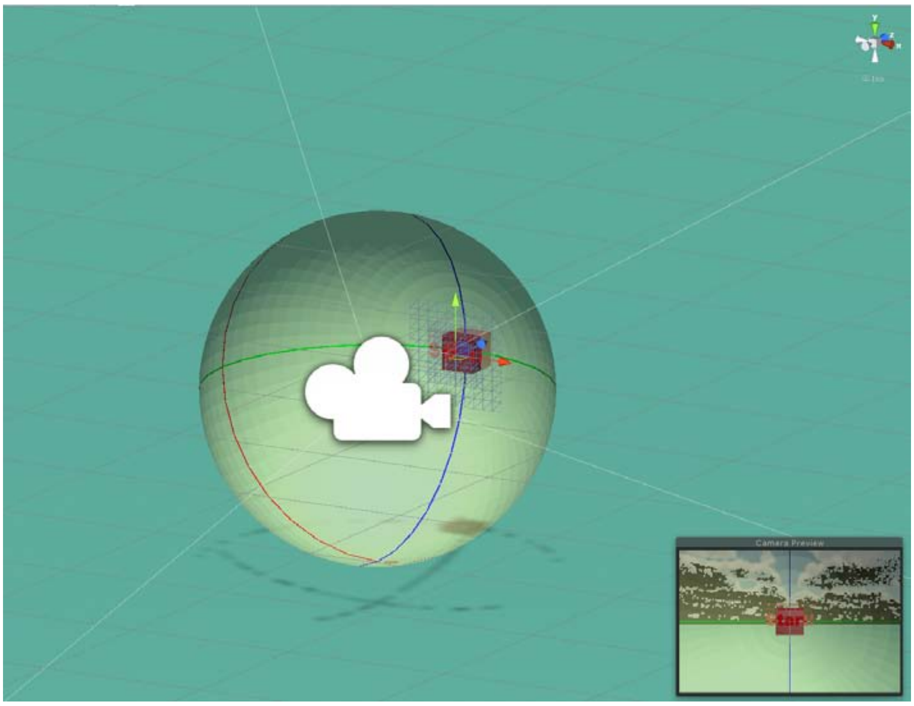


**Figure 4.2:** *Unity Experiment Scene*

to be more effective than a basic white noise burst [15]. Before the noise was used a preparation silence time of $2s$. Thus, each trial was long $5s$ (see fig. 4.3).

These stimuli are generated considering all of the possible combinations of $6$ azimuth values (from $-180°$ to $120°$ in $60°$ steps) and $4$ elevation values (from $-28.125°$ to $56.250°$ in $28.125°$ steps). It was added one last sound source positioned to the point $(0°, 90°)$. The radius of the sound sources was constant $(r = 1)$.

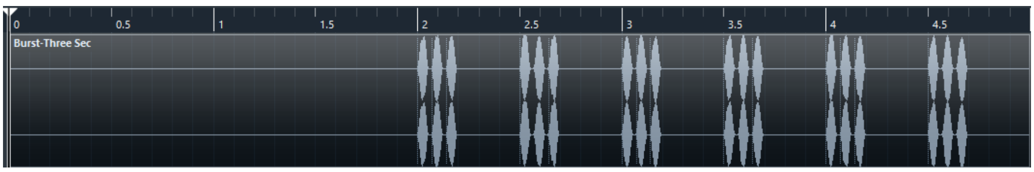



**Figure 4.3:** *5s Trail Sound*

### HRTF Conditions

For each experimental subject, the best personalized HRTF set was selected as seen in chapter 2 using the matlab application drawing pinna contours **(A)**. It was followed by an anthropometric (see 1.4.4.3 section) best match just using $d_1 + d_2 + d_4$ instead of each available measurement

**Figure 4.4:** *Three pictures of Subject-01 at three different distances*

in the CIPIC database **(C)**. This criterion was used in previous works in a similar experiment [3]. Finally, the last HRTF set **(B)** is given by default to each subject as CIPIC subject 165 (KEMAR). After that, the tested subject was collocated in the virtual reality environment, ready to follow the VR application instructions.

### 4.1.3 Procedure

#### Pinna Images Acquisition

In order to add subjects to the matlab application database, it was required to take pictures of the subjects pinna. It was important to create a stable procedure that could make possible to take every subject pinna image in the same conditions. Each subject was seated on a chair ready to follow instructions for this procedure. It was asked to the subject to center the head in a mirror placed parallel in front of him/her. The mirror had a straight vertical black line that splitted it in two halfs and it was asked to close alternatively one eye and see with the other eye if the closed one was on the vertical black line, until centered. After that, the Galaxy S7 used in the experiment was fixed to a tripod and, using three different measures,[1] from the right ear to the smartphone camera, the pinna images were taken (see fig. 4.4). Each of the three pictures had also a measuring tape near the pinna subject useful to extract the pixel-to-meter conversion.

#### HRTF selection

After the pinna acquisition, the following step was to extract the HRTF using the matlab application: the CIPIC ID that had the best match between it and the subject was selected. The

[1] 50cm, 100cm, 150cm

**Table 4.1:** *Latin Square Triads*

| | 0 | 1 | 2 |
|---|---|---|---|
| 0 | C | A | B |
| 1 | A | B | C |
| 2 | B | C | A |

first operation was to run the application, import the subject in the database with the own pinna $50cm$ image and manipulate it in order to make it easier to trace contours on it using some cropping to zoom and rotating to align the nose and the ear canal tools. The trace ear procedure was done with $N = 10$ and $k = 10$ processing the data with $M = 5$ and the pixel-to-meter tool extracted from the photo. With the pixel-to-meter conversion already acquired, it was easy to exploit an additional feature of the application in order to obtain the length to the $d_1 + d_2 + d_4$ measure.

#### Latin Square

It was likely that people would have learned how to improve their performance during the VR experience and there was the risk to see better results in the third session regardless the type of selected HRTF. That is why the order of the HRTF sets given to the subjects needed to be *randomized*. In fact, it has been decided to make a rotation using the following *latin square* criterion: the subject $i$, starting from $0$, was assigned to the $i \mod 3$ triad (see table 4.1).

In this way, each subject started the experiment with the first HRTF set, marked with the first letter in the table, switching to the second one after the end of the first session and to the third after that.

### 4.1.4 Evaluation

#### Extracted Data

After each session, a *.txt* file was stored in the smartphone with all $25$ locations data. Each row represents the real azimuth and elevation location followed by the pointed ones. To formalize how each component could change and which values can assume, we consider azimuth values $x \in [-180°, 180°[$ and elevation values $y \in [-90°, 90°]$.

### Questionnaires

After each of the three different session blocks, a questionnaire was distributed to the tested subject in order to understand the impressions about confidence and accurateness of the current session experiment.

Moreover, in order to monitor each subject physic state and check if some sort of correlation is present between it and the results, at the end of the three sessions, an ending questionnaire is given about VR experience and sickness.

All the questions in those questionnaires were in the multiple choice form and answers were given crossing only one of the possible outcomes. All answers could get an integer score from 1 to 7 except the *elevation* question which had only two possibilities (Yes/No). All these questions are mainly based on previous works [4] [16].

### Session Questionnaires

1. **Localizability**: estimating the location of the sound source was [*More Difficult 1-7 Easier]*
2. **Elevation**: did you perceive elevation? [*No 0-1 Yes]*
3. **Externalization**: was the sound source perceived inside or outside the head? *[Lower externalization 1-7 Higher externalization]*
4. **Naturalness**: how natural (close to real life) did you find the sound reproduction? *[Lower naturalness 1-7 Higher naturalness]*
5. **Presence**: to what degree did you experience a sense of ”being in space”? *[Lower 1-7 Higher]*
6. **Satisfaction**: how satisfied were you with your own performance? *[Not at all 1-7 Very much]*
7. **Confidence**: how confident were you that you pointed at the correct location of the sound source? *[Not very confident 1-7 Very confident]*

### VR Sickness Questionnaires

1. **Headtracking**: How well did the change in the area you were viewing appear to match the amount you moved your head (i.e. was the headtracking accurate)? [*Not at all 1-7 Completely]*

2. **Display**: how much did the visual display quality interfere or distract you from performing assigned tasks or required activities? [*Not at all 1-7 Very much]*

3. **Physical State**: how are you feeling today? [*Bad 1-7 Good]*

4. **Aftereffects**: did you experience any aftereffects which you think were due to your experience in the Virtual Environment? [*Not at all 1-7 Very much]*

## 4.2 Results

In this section experimental data are reported and analyzed. Starting with the questionnaires sessions and VR sickness data, it is presented the average confidence and supposed performances about the experiments. They are followed by the data analysis of the actual performances. Personal subject questionnaires and test data are presented in the *Appendix* at the end of this chapter.

### 4.2.1 Questionnaires

In this subsection, averaged data about all questionnaires are presented starting from performances of VR tests and followed by VR sickness.

#### Sessions Data

Below, tab. 4.2 shows AVG data related to what people thought about HRTF conditions in average and tab. 4.3 shows AVG data related to what people thought about each of they three sessions.

Avg ABC

| | Local | Elev | Ext | Natu | Presence | Satis | Confi |
|---|---|---|---|---|---|---|---|
| A | 4,07 | 0,73 | 4,33 | 4,07 | 4,27 | 4,00 | 3,73 |
| B | 4,20 | 0,40 | 4,60 | 3,87 | 4,40 | 3,80 | 3,47 |
| C | 4,87 | 0,87 | 4,47 | 4,20 | 4,60 | 4,67 | 4,33 |

**Table 4.2:** *Questionnaires of the three VR sessions (Average on HRTF condition)*

| Avg Learning | | | | | | | |
|---|---|---|---|---|---|---|---|
| | Local | Elev | Ext | Natu | Presence | Satis | Confi |
| 1° | 4,467 | 0,533 | 4,067 | 3,800 | 4,333 | 3,600 | 3,667 |
| 2° | 4,333 | 0,667 | 5,067 | 4,067 | 4,333 | 4,533 | 4,000 |
| 3° | 4,333 | 0,800 | 4,267 | 4,267 | 4,600 | 4,333 | 3,867 |

**Table 4.3:** *Questionnaires of the three VR sessions (Average on the number of the session)*

#### VR sickness Data

In tab. 4.4 the overall view of VR sickness data can be seen. In average, all subjects thought that the headtracking was accurate, the display did not interfere with the experiment, they felt good that day and did not have any aftereffect due to the VR experience.

| Headtracking | Display | How Feel | Aftereffects |
|---|---|---|---|
| 5,87 | 2,67 | 5,67 | 2,13 |

**Table 4.4:** *Averaged results about VR sickness questionnaires considering all subjects*

### 4.2.2 VR Tests

Creating an Excel file organized in 15 sheets, one per subject, average and standard deviation calculations are done for the azimuth and elevation error components. After that, they are copied and pasted into a new sheet to have a better compact view of data and results[2].

#### Azimuth and Elevation Errors

With regard to the azimuth error, it is has been decided to take the minimum error between the perceived azimuth location and the real or the cone of confusion one at the same elevation:

$$X_{Err_1} = \begin{cases} ||x_{real} - x_{perc}| - 360°| & \text{if } |x_{real} - x_{perc}| > 180° \\ |x_{real} - x_{perc}| & \text{otherwise} \end{cases} \tag{4.1}$$

$$X_{Err_2} = \begin{cases} ||Cone(x_{perc}) - x_{real}| - 360°| & \text{if } |Cone(x_{perc}) - x_{real}| > 180° \\ |Cone(x_{perc}) - x_{real}| & \text{otherwise} \end{cases} \tag{4.2}$$

[2]To simplify each table view, the degree simbol to each elevation and azimuth number has been removed

$$X_{Err} = \min\{X_{Err_1}, X_{Err_2}\} \tag{4.3}$$

where $Cone(x)$ gives as output the other azimuth value in the cone of confusion that has the same absolute angular difference as $x$ with respect to the origin of the hemisphere ($0°$ for the front one, $-180°$ or $180°$ for the back one). $x_{real}$ is the real source azimuth location and the $x_{perc}$ is the perceived azimuth value pointed by the subject.

The elevation error is calculated as the absolute difference between perceived elevation and real one:

$$Y_{Err} = |y_{real} - y_{perc}| \tag{4.4}$$

### Main Results

Starting with the raw data in the 15 tables, the azimuth and elevation error is calculated using the equations 4.3 4.4 above. Figure 4.5 shows all session AVG results for each HRTF condition and subject. Figure 4.6 shows all session StD results for each HRTF condition and subject.

Fig. 4.5 shows the linear regression using all the raw data of the perceived elevation for each HRTF condition. The closer the slope is to 1, the more the elevation was perceived in a correctly way.

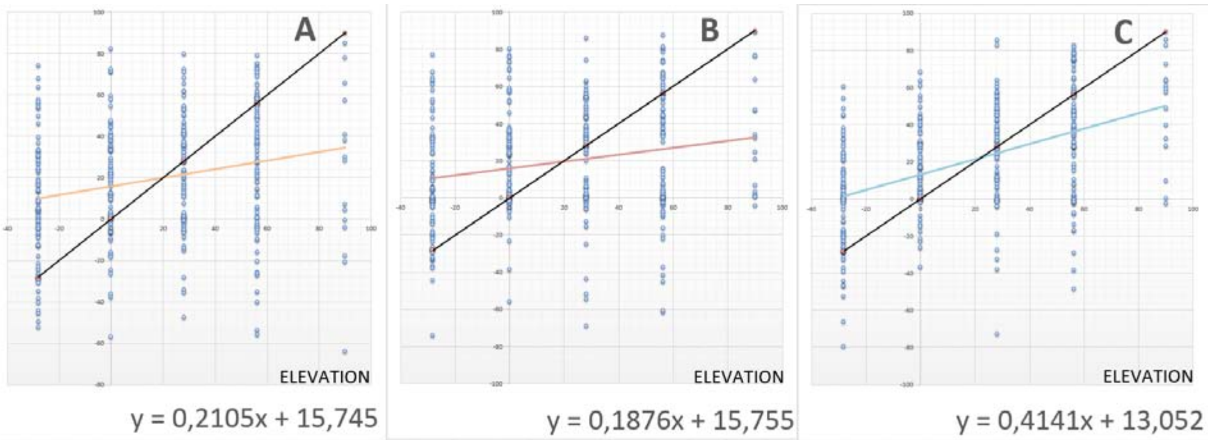



**Figure 4.5:** *Linear regression on all raw elevation data for each HRTF condition (all subjects).*

### Learning curve

The first important issue was to check how much people would have learnt from the VR experiment, session after session. Questionnaires gave the idea that there was not a fast growing

AllSubjects

| | AVG Azimuth | | | AVG Elevation | | |
|---|---|---|---|---|---|---|
| | A | B | C | A | B | C |
| 1 | 10,13 | 8,06 | 5,56 | 21,56 | 21,83 | 21,18 |
| 2 | 9,13 | 13,76 | 14,51 | 23,46 | 23,91 | 17,89 |
| 3 | 21,14 | 83,04 | 10,96 | 41,04 | 35,34 | 23,60 |
| 4 | 5,90 | 9,08 | 8,33 | 30,41 | 34,13 | 19,90 |
| 5 | 24,48 | 28,24 | 21,28 | 39,50 | 30,87 | 29,52 |
| 6 | 20,63 | 16,82 | 12,73 | 20,43 | 28,55 | 17,88 |
| 7 | 23,44 | 16,42 | 10,06 | 23,44 | 27,02 | 27,12 |
| 8 | 19,55 | 17,51 | 17,14 | 29,77 | 38,74 | 28,17 |
| 9 | 11,73 | 15,42 | 11,69 | 30,86 | 33,25 | 26,83 |
| 10 | 15,70 | 25,36 | 32,14 | 37,17 | 44,08 | 34,30 |
| 11 | 15,87 | 12,64 | 25,49 | 35,02 | 32,48 | 37,47 |
| 12 | 9,88 | 10,69 | 13,84 | 28,77 | 32,33 | 20,21 |
| 13 | 10,44 | 15,10 | 12,26 | 27,34 | 25,92 | 25,56 |
| 14 | 21,04 | 16,80 | 10,20 | 31,30 | 40,36 | 28,71 |
| 15 | 14,67 | 17,61 | 11,61 | 24,18 | 30,15 | 27,43 |
| AVG | 15,58 | 20,44 | 14,52 | 29,62 | 31,93 | 25,72 |

**Table 4.5:** *AVG azimuth error (yellow) and AVG elevation (green) of the VR experiment test for each subject and HRTF condition (A,B,C). Rows represent the AVG error done for each condition and subject in the experiment.*

learning curve except for the elevation perception. (see column *Elev* in tab. 4.3). Since the questionnaires did not give critical subject results (i.e. someone that could not understand at all the correct sound localization or someone that was too much affected by VR sickness), an additional analysis with the actual data is needed in order to understand how much learning influenced perception in these experiments.

Doing the average on the mean values of each error session and organized them in a table and follow the order of the latin square criterion (see tab. 4.5), it was easy to extract table 4.8. The errors values are quite close to each other thus there was not an average quality improvement of the score incrementing the number session even if this fact was not coherent with the elevation data already seen in the questionnaires (tab. 4.3).

| AllSubjects | StD Azimuth | | | StD Elevation | | |
|---|---|---|---|---|---|---|
| | A | B | C | A | B | C |
| 1 | 14,46 | 17,81 | 5,70 | 16,74 | 21,50 | 15,84 |
| 2 | 7,81 | 17,21 | 17,09 | 22,58 | 23,43 | 15,07 |
| 3 | 20,80 | 55,16 | 13,17 | 28,02 | 21,81 | 17,92 |
| 4 | 6,07 | 6,61 | 8,03 | 34,15 | 24,03 | 20,96 |
| 5 | 21,33 | 22,81 | 26,33 | 30,33 | 23,58 | 25,84 |
| 6 | 22,67 | 12,81 | 10,57 | 18,15 | 26,06 | 11,84 |
| 7 | 17,18 | 16,82 | 13,85 | 17,18 | 24,94 | 20,37 |
| 8 | 15,65 | 15,09 | 13,64 | 24,39 | 26,93 | 21,06 |
| 9 | 11,60 | 14,23 | 8,74 | 19,89 | 24,31 | 19,10 |
| 10 | 14,08 | 20,34 | 22,53 | 28,23 | 32,40 | 22,76 |
| 11 | 19,71 | 12,83 | 29,72 | 26,74 | 29,26 | 30,93 |
| 12 | 9,56 | 14,62 | 12,29 | 20,94 | 24,40 | 20,01 |
| 13 | 6,69 | 18,90 | 9,69 | 16,51 | 18,25 | 18,35 |
| 14 | 16,23 | 18,81 | 10,34 | 25,59 | 31,10 | 21,41 |
| 15 | 15,48 | 21,10 | 13,78 | 18,96 | 21,91 | 17,97 |
| AVG | 14,62 | 19,01 | 14,37 | 23,23 | 24,93 | 19,96 |

**Table 4.6:** *StD azimuth error (yellow) and StD elevation error (green) of the VR experiment test for each subject and HRTF condition (A,B,C). Rows represent the StD for each condition and subject in the experiment.*

### Filtering Data

An overall analysis of the results allows to discard some *outliers*, i.e. subjects who behaved very differently average results. In particular subject-03 can be removed due to his extremely high number of front-back confusion errors in the (B) condition. Moreover, in particular two subjects (05,10) did not perceive elevation at all in every HRTF condition[3] (see fig. 5.5-5.10).

Eliminating these subjects, new tables for each condition can be made with all the subjects that had not those results.

Tab. 4.11 represents the new slope and front-back confusion tables among the selected subjects.

[3] All personal regression graphs can be seen in the *Appendix* chapter.

| | FrontBackALL | | | | SlopesALL | | |
|---|---|---|---|---|---|---|---|
| | A | B | C | | A | B | C |
| 1 | 1 | 4 | 1 | 1 | 0,404 | 0,501 | 0,660 |
| 2 | 1 | 0 | 1 | 2 | 0,245 | 0,254 | 0,547 |
| 3 | 1 | 22 | 2 | 3 | 0,234 | 0,129 | 0,560 |
| 4 | 1 | 1 | 1 | 4 | 0,136 | 0,367 | 0,488 |
| 5 | 6 | 5 | 5 | 5 | 0,080 | 0,020 | 0,191 |
| 6 | 3 | 3 | 1 | 6 | 0,462 | 0,324 | 0,817 |
| 7 | 5 | 3 | 1 | 7 | 0,046 | 0,219 | 0,233 |
| 8 | 1 | 2 | 2 | 8 | 0,281 | 0,011 | 0,276 |
| 9 | 3 | 4 | 1 | 9 | 0,190 | 0,000 | 0,291 |
| 10 | 3 | 7 | 5 | 10 | -0,128 | -0,106 | -0,012 |
| 11 | 2 | 3 | 3 | 11 | 0,183 | 0,414 | 0,382 |
| 12 | 0 | 0 | 2 | 12 | 0,200 | 0,013 | 0,344 |
| 13 | 1 | 1 | 0 | 13 | 0,288 | 0,301 | 0,327 |
| 14 | 3 | 2 | 3 | 14 | 0,210 | 0,320 | 0,670 |
| 15 | 3 | 2 | 3 | 15 | 0,480 | 0,016 | 0,420 |
| AVG | 2,27 | 3,93 | 2,07 | AVG | 0,22 | 0,19 | 0,41 |

**Table 4.7:** *Front-back confusion occurrences for each HRTF condition (Left). Linear regression slopes for elevation data for each HRTF condition (Right).*

| AVG ERR. AZIMUTH (ALL SUBJECTS-Learning) | | |
|---|---|---|
| 1 | 2 | 3 |
| 18,88 | 13,46 | 15,04 |
| **AVG ERR. ELEVATION (ALL SUBJECTS-Learning)** | | |
| 1 | 2 | 3 |
| 27,92 | 26,39 | 27,50 |

**Table 4.8:** *AVG azimuth error (yellow) and AVG elevation (green) of the VR experiment test for each subject and number of session*

### T-Test

Paired t-test were done for each couple of HRTF conditions. Tables below (4.12-4.13-4.14-4.15-4.16-4.14) show the six t-test results. Following the *Bonferroni correction* for multiple t-tests, the relevant value of $p$ will be $0.025$.

With regard to **azimuth**, no pair condition was found statistically significant different, since each $p$ was higher than $0.025$.

With regard to **elevation**, **(A)** and **(B)** conditions were found statistically significant different

| | NO 3-5-10 | | | | | |
|---|---|---|---|---|---|---|
| | AVG Azi | | | AVG Ele | | |
| | A | B | C | A | B | C |
| 1 | 10,13 | 8,06 | 5,56 | 21,56 | 21,83 | 21,18 |
| 2 | 9,13 | 13,76 | 14,51 | 23,46 | 23,91 | 17,89 |
| 4 | 5,90 | 9,08 | 8,33 | 30,41 | 34,13 | 19,90 |
| 6 | 20,63 | 16,82 | 12,73 | 20,43 | 28,55 | 17,88 |
| 7 | 23,44 | 16,42 | 10,06 | 23,44 | 27,02 | 27,12 |
| 8 | 19,55 | 17,51 | 17,14 | 29,77 | 38,74 | 28,17 |
| 9 | 11,73 | 15,42 | 11,69 | 30,86 | 33,25 | 26,83 |
| 11 | 15,87 | 12,64 | 25,49 | 35,02 | 32,48 | 37,47 |
| 12 | 9,88 | 10,69 | 13,84 | 28,77 | 32,33 | 20,21 |
| 13 | 10,44 | 15,10 | 12,26 | 27,34 | 25,92 | 25,56 |
| 14 | 21,04 | 16,80 | 10,20 | 31,30 | 40,36 | 28,71 |
| 15 | 14,67 | 17,61 | 11,61 | 24,18 | 30,15 | 27,43 |
| AVG | 14,37 | 14,16 | 12,79 | 27,21 | 30,72 | 24,86 |

**Table 4.9:** *AVG azimuth error (yellow) and AVG elevation (green) of the VR experiment test for each subject and HRTF condition (A,B,C). Rows represent the AVG error done for each condition and subject in the experiment. Without subjects 03,05,10.*

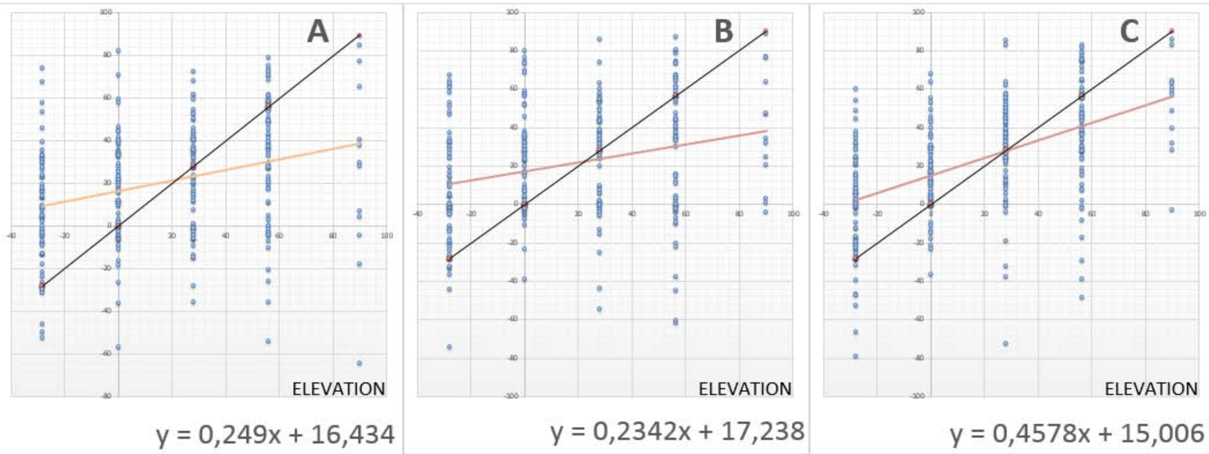



**Figure 4.6:** *Linear regression on all raw elevation data for each HRTF condition. Without subjects 03,05,10.*

having $p = 0.005 < 0.025$.

With regard to **(B)** and **(C)** conditions a statistically significant difference was found having $p = 0.0034 < 0.025$.

With regard to **(A)** and **(C)** conditions no statistically significant difference was found since

NO 3-5-10

| | StD Azimuth | | | StD Elevation | | |
|---|---|---|---|---|---|---|
| | A | B | C | A | B | C |
| 1 | 14,46 | 17,81 | 5,70 | 16,74 | 21,50 | 15,84 |
| 2 | 7,81 | 17,21 | 17,09 | 22,58 | 23,43 | 15,07 |
| 4 | 6,07 | 6,61 | 8,03 | 34,15 | 24,03 | 20,96 |
| 6 | 22,67 | 12,81 | 10,57 | 18,15 | 26,06 | 11,84 |
| 7 | 17,18 | 16,82 | 13,85 | 17,18 | 24,94 | 20,37 |
| 8 | 15,65 | 15,09 | 13,64 | 24,39 | 26,93 | 21,06 |
| 9 | 11,60 | 14,23 | 8,74 | 19,89 | 24,31 | 19,10 |
| 11 | 19,71 | 12,83 | 29,72 | 26,74 | 29,26 | 30,93 |
| 12 | 9,56 | 14,62 | 12,29 | 20,94 | 24,40 | 20,01 |
| 13 | 6,69 | 18,90 | 9,69 | 16,51 | 18,25 | 18,35 |
| 14 | 16,23 | 18,81 | 10,34 | 25,59 | 31,10 | 21,41 |
| 15 | 15,48 | 21,10 | 13,78 | 18,96 | 21,91 | 17,97 |
| AVG | 13,59 | 15,57 | 12,79 | 21,82 | 24,68 | 19,41 |

**Table 4.10:** *StD azimuth error (yellow) and StD elevation error (green) of the VR experiment test for each subject and HRTF condition (A,B,C). Rows represent the StD for each condition and subject in the experiment. Without subjects 03,05,10.*

| | FrontBack | | |
|---|---|---|---|
| | A | B | C |
| 1 | 1 | 4 | 1 |
| 2 | 1 | 0 | 1 |
| 4 | 1 | 1 | 1 |
| 6 | 3 | 3 | 1 |
| 7 | 5 | 3 | 1 |
| 8 | 1 | 2 | 2 |
| 9 | 3 | 4 | 1 |
| 11 | 2 | 3 | 3 |
| 12 | 0 | 0 | 2 |
| 13 | 1 | 1 | 0 |
| 14 | 3 | 2 | 3 |
| 15 | 3 | 2 | 3 |
| AVG | 2,00 | 2,08 | 1,58 |

| | Slopes | | |
|---|---|---|---|
| | A | B | C |
| 1 | 0,404 | 0,501 | 0,660 |
| 2 | 0,245 | 0,254 | 0,547 |
| 4 | 0,136 | 0,367 | 0,488 |
| 6 | 0,462 | 0,324 | 0,817 |
| 7 | 0,046 | 0,219 | 0,233 |
| 8 | 0,281 | 0,011 | 0,276 |
| 9 | 0,190 | 0,000 | 0,291 |
| 11 | 0,183 | 0,414 | 0,382 |
| 12 | 0,200 | 0,013 | 0,344 |
| 13 | 0,288 | 0,301 | 0,327 |
| 14 | 0,210 | 0,320 | 0,670 |
| 15 | 0,480 | 0,016 | 0,420 |
| AVG | 0,26 | 0,23 | 0,45 |

**Table 4.11:** *Front-back confusion occurrences for each HRTF condition (Left). Linear regression slopes for elevation data for each HRTF condition (Right). Without subjects 03,05,10.*

$p = 0.046 > 0.025$.

| NO 3-5-10 | | |
|---|---|---|
| | *A* | *B* |
| AVG | 14,37 | 14,16 |
| Variance | 32,26 | 11,12 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,722115 | |
| Stat t | 0,181123 | |
| P(T<=t) one tail | 0,429782 | |

**Table 4.12:** *Azimuth errors paired T-Test for the HRTF conditions AB. Without subjects 03,05,10*

| NO 3-5-10 | | |
|---|---|---|
| | *A* | *C* |
| AVG | 14,37 | 12,79 |
| Variance | 32,26 | 24,85 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,188078 | |
| Stat t | 0,803899 | |
| P(T<=t) one tail | 0,219242 | |

**Table 4.13:** *Azimuth errors paired T-Test for the HRTF conditions AC. Without subjects 03,05,10*

| NO 3-5-10 | | |
|---|---|---|
| | *B* | *C* |
| AVG | 14,16 | 12,79 |
| Variance | 11,12 | 24,85 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,21124171 | |
| Stat t | 0,88369988 | |
| P(T<=t) one tail | 0,19788094 | |

**Table 4.14:** *Azimuth errors paired T-Test for the HRTF conditions BC. Without subjects 03,05,10*

| NO 3-5-10 | | |
|---|---|---|
| | *A* | *B* |
| AVG | 27,21 | 30,72 |
| Variance | 20,52 | 31,64 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,720436 | |
| Stat t | -3,09529 | |
| P(T<=t) one tail | 0,005095 | |

**Table 4.15:** *Elevation errors paired T-Test for the HRTF conditions AB. Without subjects 03,05,10*

| NO 3-5-10 | | |
|---|---|---|
| | *A* | *C* |
| AVG | 27,21 | 24,86 |
| Variance | 20,52 | 32,56 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,650443 | |
| Stat t | 1,844141 | |
| P(T<=t) one tail | 0,046121 | |

**Table 4.16:** *Elevation errors paired T-Test for the HRTF conditions AC. Without subjects 03,05,10*

| NO 3-5-10 | | |
|---|---|---|
| | *B* | *C* |
| AVG | 30,72 | 24,86 |
| Variance | 31,64 | 32,56 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,42078025 | |
| Stat t | 3,328633 | |
| P(T<=t) one tail | 0,00336377 | |

**Table 4.17:** *Elevation errors paired T-Test for the HRTF conditions BC. Without subjects 03,05,10*

## 4.3 Conclusions

All questionnaires were coherent with the results except for the learning elevation curve. People were more confidence with the **(C)** condition followed by the **(A)** and (B) ones and the AVG results confirmed that order of preference.

#### Azimuth

Since each t-test gave $p$ higher than the requested value for statistically significant difference, we can conclude that each of the three condition performed in the same way with regard to the azimuth component. This is according to the fact that all the conditions were based on pinna and it involves only elevations performances.

#### Elevation

Comparing HRTF elevation results of the selected subjects, it can be possible to say that the **C** condition performed better in elevation, followed by the **A** condition and the **B** one. In average the elevation error results are very similar but the regressions on all data gave a result of an almost double slope of **C** with respect to **A** and **B**.

On the other hand, t-tests gave the result that both **(A)** and **(C)** conditions are different statistically from the **(B)** one but not different from each other.

We can conclude that **(A)** and **(C)** conditions are statistically better than the KEMAR generic function.

### 4.3.1 Future works

Future applications can be done in order to improve the **(A)** method. A better weighted metric can be found to compare each of the three output classifications in addition to find some criterions to select subjects in a pre-processed way to guarantee high results or avoid bad ones. In a related work submitted for publication [9], the validation of our approach is carried out improving in the meantime the way to score CIPIC subjects, aiming to better results.

# Chapter 5

# Appendix

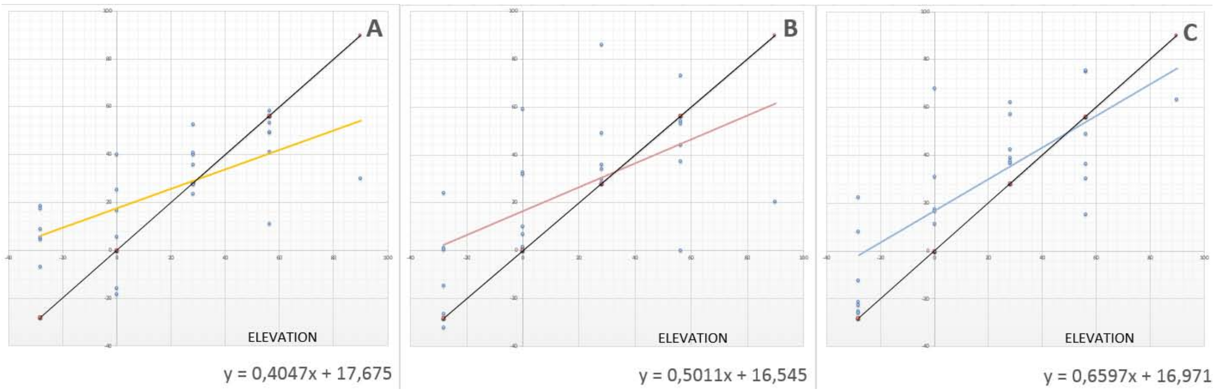



**Figure 5.1:** *Linear regression for elevation data for subject 01 for each HRTF condition*

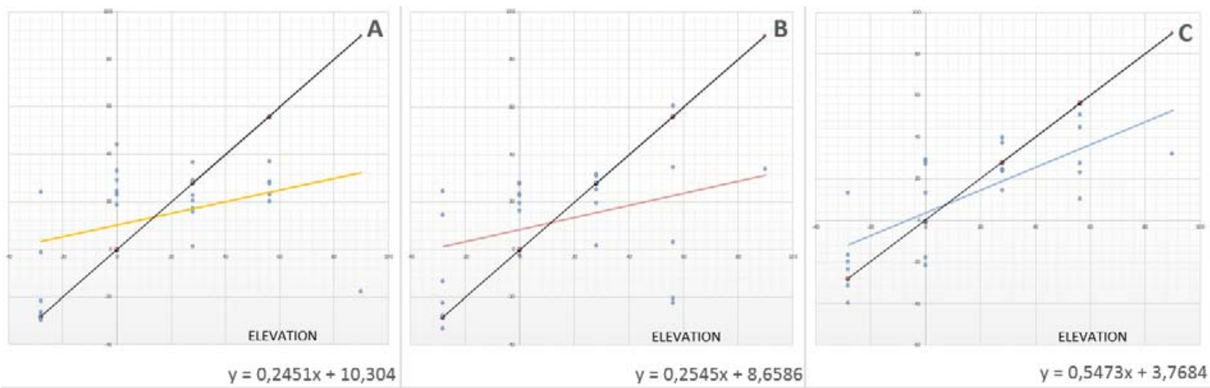



**Figure 5.2:** *Linear regression for elevation data for subject 02 for each HRTF condition*

| | HRTF | Local | Elev | Ext | Natu | Presence | Satis | Confi |
|---|---|---|---|---|---|---|---|---|
| 1 | A | 1 | 0 | 1 | 3 | 4 | 2 | 2 |
| 1 | B | 2 | 0 | 2 | 3 | 4 | 2 | 2 |
| 1 | C | 2 | 0 | 2 | 4 | 5 | 2 | 2 |
| | AVG | 1,67 | 0,00 | 1,67 | 3,33 | 4,33 | 2,00 | 2,00 |
| | Std | 0,58 | 0,00 | 0,58 | 0,58 | 0,58 | 0,00 | 0,00 |
| 2 | A | 6 | 1 | 3 | 4 | 6 | 5 | 5 |
| 2 | B | 3 | 1 | 4 | 3 | 4 | 3 | 4 |
| 2 | C | 5 | 1 | 5 | 4 | 5 | 5 | 4 |
| | AVG | 4,67 | 1,00 | 4,00 | 3,67 | 5,00 | 4,33 | 4,33 |
| | Std | 1,53 | 0,00 | 1,00 | 0,58 | 1,00 | 1,15 | 0,58 |
| 3 | A | 4 | 1 | 5 | 4 | 5 | 5 | 5 |
| 3 | B | 5 | 0 | 5 | 3 | 3 | 3 | 3 |
| 3 | C | 5 | 1 | 5 | 5 | 5 | 6 | 5 |
| | AVG | 4,67 | 0,67 | 5,00 | 4,00 | 4,33 | 4,67 | 4,33 |
| | Std | 0,58 | 0,58 | 0,00 | 1,00 | 1,15 | 1,53 | 1,15 |
| 4 | A | 3 | 0 | 3 | 3 | 2 | 4 | 3 |
| 4 | B | 5 | 1 | 5 | 5 | 4 | 5 | 5 |
| 4 | C | 3 | 1 | 2 | 3 | 3 | 2 | 3 |
| | AVG | 3,67 | 0,67 | 3,33 | 3,67 | 3,00 | 3,67 | 3,67 |
| | Std | 1,15 | 0,58 | 1,53 | 1,15 | 1,00 | 1,53 | 1,15 |
| 5 | A | 2 | 1 | 6 | 3 | 1 | 2 | 2 |
| 5 | B | 3 | 0 | 7 | 4 | 2 | 3 | 3 |
| 5 | C | 6 | 1 | 3 | 6 | 2 | 5 | 5 |
| | AVG | 3,67 | 0,67 | 5,33 | 4,33 | 1,67 | 3,33 | 3,33 |
| | Std | 2,08 | 0,58 | 2,08 | 1,53 | 0,58 | 1,53 | 1,53 |
| 6 | A | 6 | 1 | 3 | 5 | 6 | 6 | 6 |
| 6 | B | 4 | 0 | 4 | 2 | 3 | 2 | 3 |
| 6 | C | 5 | 1 | 5 | 4 | 5 | 4 | 4 |
| | AVG | 5,00 | 0,67 | 4,00 | 3,67 | 4,67 | 4,00 | 4,33 |
| | Std | 1,00 | 0,58 | 1,00 | 1,53 | 1,53 | 2,00 | 1,53 |
| 7 | A | 6 | 1 | 6 | 5 | 5 | 6 | 5 |
| 7 | B | 4 | 0 | 2 | 4 | 5 | 4 | 4 |
| 7 | C | 4 | 1 | 3 | 4 | 5 | 4 | 4 |
| | AVG | 4,67 | 0,67 | 3,67 | 4,33 | 5,00 | 4,67 | 4,33 |
| | Std | 1,15 | 0,58 | 2,08 | 0,58 | 0,00 | 1,15 | 0,58 |

**Table 5.1:** *Questionnaires of the three VR sessions (First part)*

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | AVG | 4,67 | 0,67 | 3,67 | 4,33 | 5,00 | 4,67 | 4,33 |
| | Std | 1,15 | 0,58 | 2,08 | 0,58 | 0,00 | 1,15 | 0,58 |
| 8 | A | 5 | 1 | 5 | 5 | 2 | 4 | 4 |
| 8 | B | 4 | 1 | 4 | 5 | 2 | 4 | 3 |
| 8 | C | 6 | 1 | 5 | 5 | 3 | 5 | 5 |
| | AVG | 5,00 | 1,00 | 4,67 | 5,00 | 2,33 | 4,33 | 4,00 |
| | Std | 1,00 | 0,00 | 0,58 | 0,00 | 0,58 | 0,58 | 1,00 |
| 9 | A | 4 | 0 | 5 | 4 | 4 | 4 | 5 |
| 9 | B | 6 | 0 | 2 | 6 | 6 | 4 | 4 |
| 9 | C | 5 | 1 | 6 | 3 | 3 | 6 | 5 |
| | AVG | 5,00 | 0,33 | 4,33 | 4,33 | 4,33 | 4,67 | 4,67 |
| | Std | 1,00 | 0,58 | 2,08 | 1,53 | 1,53 | 1,15 | 0,58 |
| 10 | A | 4 | 0 | 6 | 3 | 3 | 4 | 3 |
| 10 | B | 4 | 1 | 5 | 2 | 4 | 3 | 2 |
| 10 | C | 5 | 0 | 5 | 2 | 3 | 3 | 3 |
| | AVG | 4,33 | 0,33 | 5,33 | 2,33 | 3,33 | 3,33 | 2,67 |
| | Std | 0,58 | 0,58 | 0,58 | 0,58 | 0,58 | 0,58 | 0,58 |
| 11 | A | 2 | 1 | 3 | 3 | 4 | 2 | 2 |
| 11 | B | 4 | 0 | 6 | 2 | 5 | 4 | 3 |
| 11 | C | 4 | 1 | 5 | 3 | 5 | 5 | 4 |
| | AVG | 3,33 | 0,67 | 4,67 | 2,67 | 4,67 | 3,67 | 3,00 |
| | Std | 1,15 | 0,58 | 1,53 | 0,58 | 0,58 | 1,53 | 1,00 |
| 12 | A | 4 | 1 | 3 | 3 | 4 | 4 | 3 |
| 12 | B | 6 | 0 | 7 | 6 | 6 | 5 | 4 |
| 12 | C | 6 | 1 | 7 | 6 | 7 | 6 | 6 |
| | AVG | 5,33 | 0,67 | 5,67 | 5,00 | 5,67 | 5,00 | 4,33 |
| | Std | 1,15 | 0,58 | 2,31 | 1,73 | 1,53 | 1,00 | 1,53 |
| 13 | A | 5 | 1 | 6 | 7 | 7 | 5 | 4 |
| 13 | B | 3 | 1 | 7 | 7 | 7 | 5 | 2 |
| 13 | C | 7 | 1 | 5 | 6 | 7 | 7 | 6 |
| | AVG | 5,00 | 1,00 | 6,00 | 6,67 | 7,00 | 5,67 | 4,00 |
| | Std | 2,00 | 0,00 | 1,00 | 0,58 | 0,00 | 1,15 | 2,00 |
| 14 | A | 5 | 1 | 5 | 4 | 5 | 4 | 5 |
| 14 | B | 5 | 1 | 5 | 4 | 5 | 5 | 5 |
| 14 | C | 4 | 1 | 4 | 4 | 5 | 4 | 4 |
| | AVG | 4,67 | 1,00 | 4,67 | 4,00 | 5,00 | 4,33 | 4,67 |
| | Std | 0,58 | 0,00 | 0,58 | 0,00 | 0,00 | 0,58 | 0,58 |
| 15 | A | 4 | 1 | 5 | 5 | 6 | 3 | 2 |
| 15 | B | 5 | 0 | 4 | 2 | 6 | 5 | 5 |
| 15 | C | 6 | 1 | 5 | 4 | 6 | 6 | 5 |
| | AVG | 5,00 | 0,67 | 4,67 | 3,67 | 6,00 | 4,67 | 4,00 |
| | Std | 1,00 | 0,58 | 0,58 | 1,53 | 0,00 | 1,53 | 1,73 |

**Table 5.2:** *Questionnaires of the three VR sessions (Second part)*

| | Headtracking | Display | How Feel | Aftereffects |
|---|---|---|---|---|
| 1 | 7 | 1 | 4 | 1 |
| 2 | 6 | 5 | 6 | 2 |
| 3 | 7 | 1 | 7 | 1 |
| 4 | 6 | 2 | 7 | 1 |
| 5 | 5 | 2 | 6 | 1 |
| 6 | 6 | 3 | 7 | 2 |
| 7 | 6 | 4 | 6 | 4 |
| 8 | 3 | 6 | 5 | 2 |
| 9 | 4 | 1 | 6 | 3 |
| 10 | 7 | 3 | 7 | 1 |
| 11 | 6 | 3 | 3 | 3 |
| 12 | 6 | 2 | 7 | 3 |
| 13 | 7 | 1 | 7 | 1 |
| 14 | 5 | 3 | 3 | 3 |
| 15 | 7 | 3 | 4 | 4 |

**Table 5.3:** *Subjects answers for VR sickness questionnaires*

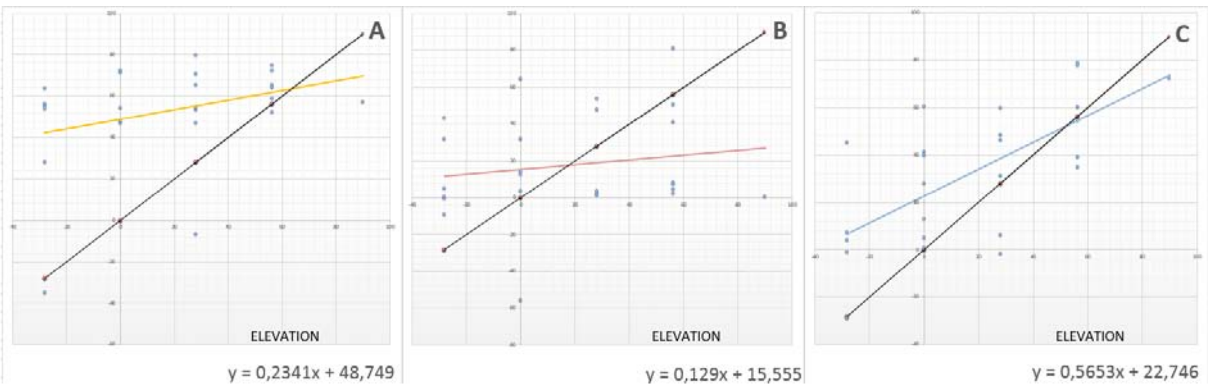



**Figure 5.3:** *Linear regression for elevation data for subject 03 for each HRTF condition*

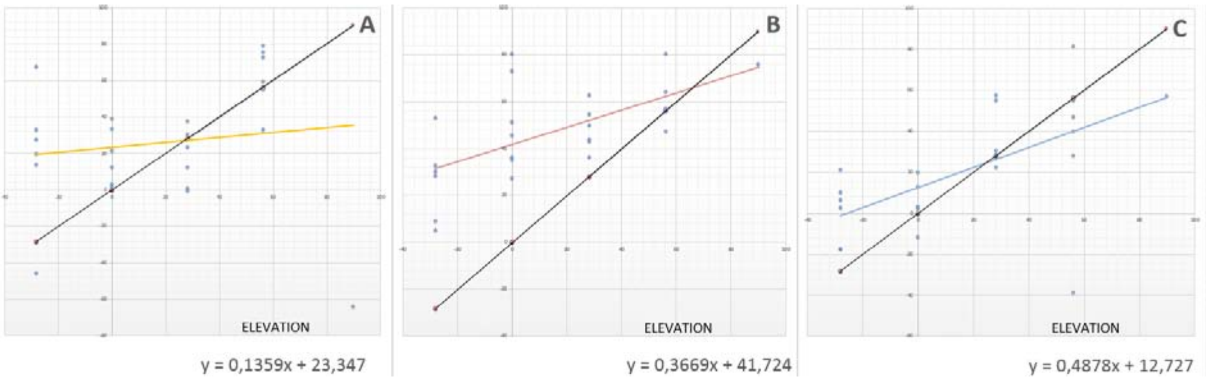



**Figure 5.4:** *Linear regression for elevation data for subject 04 for each HRTF condition*

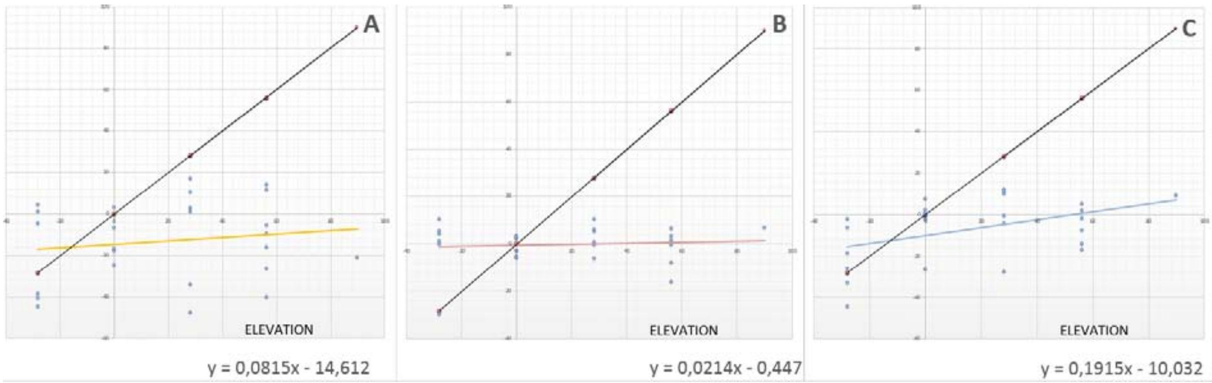



**Figure 5.5:** *Linear regression for elevation data for subject 05 for each HRTF condition*

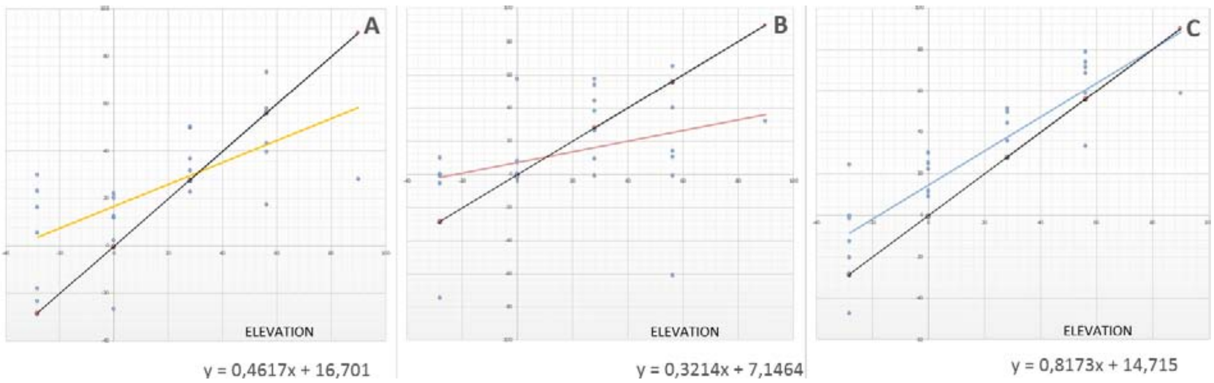



**Figure 5.6:** *Linear regression for elevation data for subject 06 for each HRTF condition*

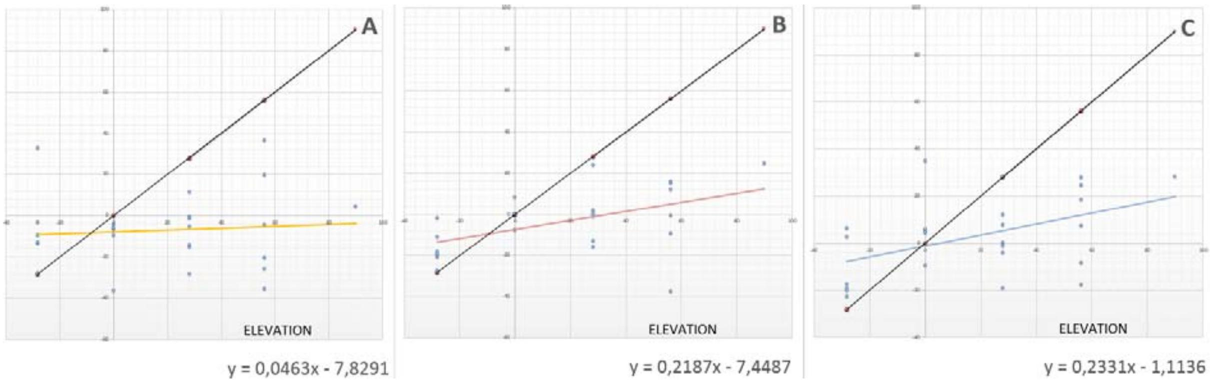



**Figure 5.7:** *Linear regression for elevation data for subject 07 for each HRTF condition*

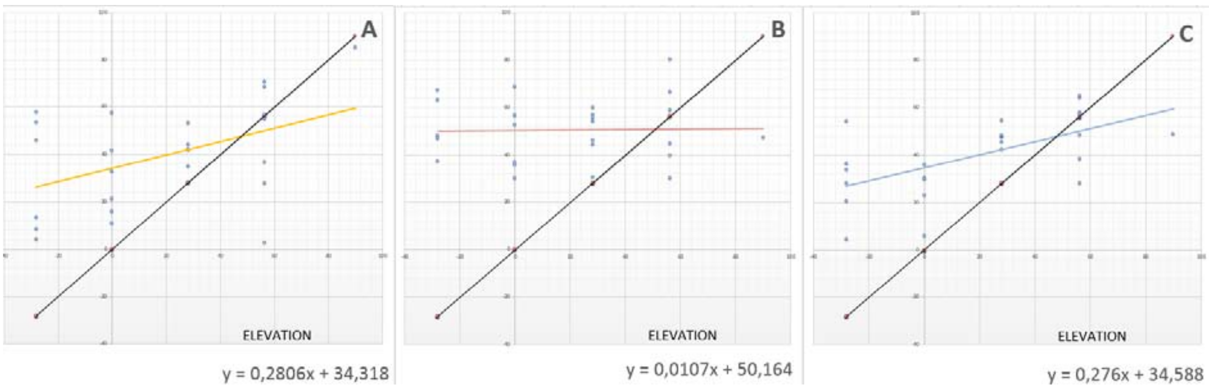



**Figure 5.8:** *Linear regression for elevation data for subject 08 for each HRTF condition*

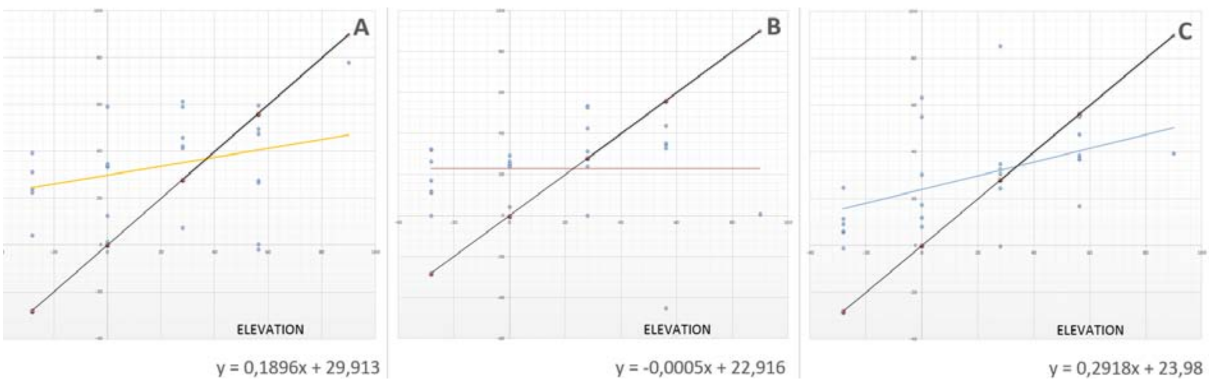



**Figure 5.9:** *Linear regression for elevation data for subject 09 for each HRTF condition*

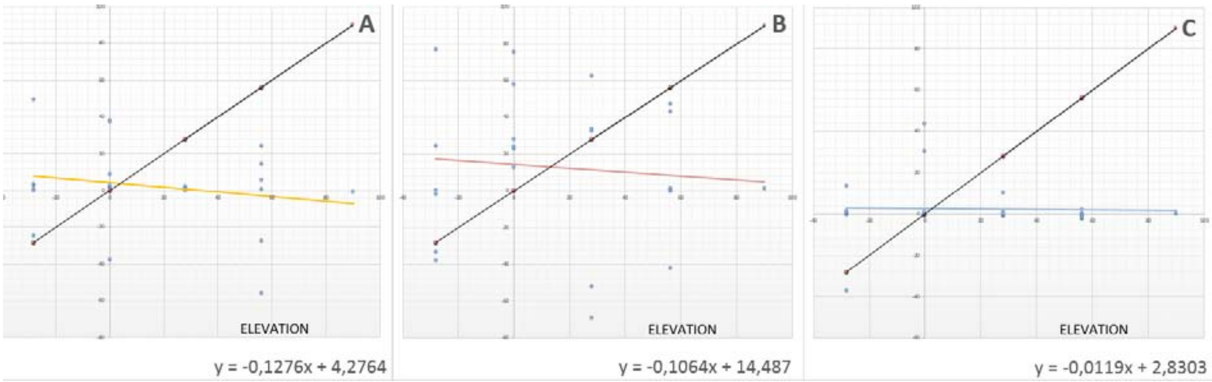



**Figure 5.10:** *Linear regression for elevation data for subject 010 for each HRTF condition*

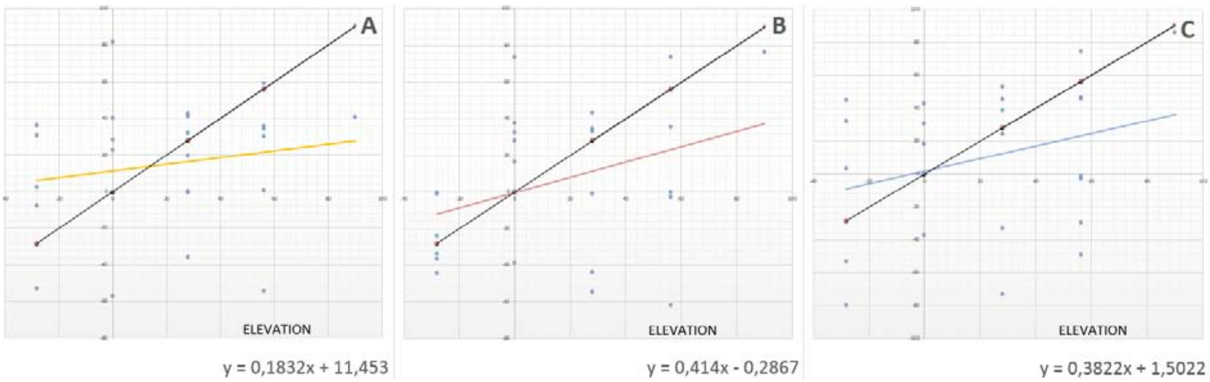



**Figure 5.11:** *Linear regression for elevation data for subject 011 for each HRTF condition*

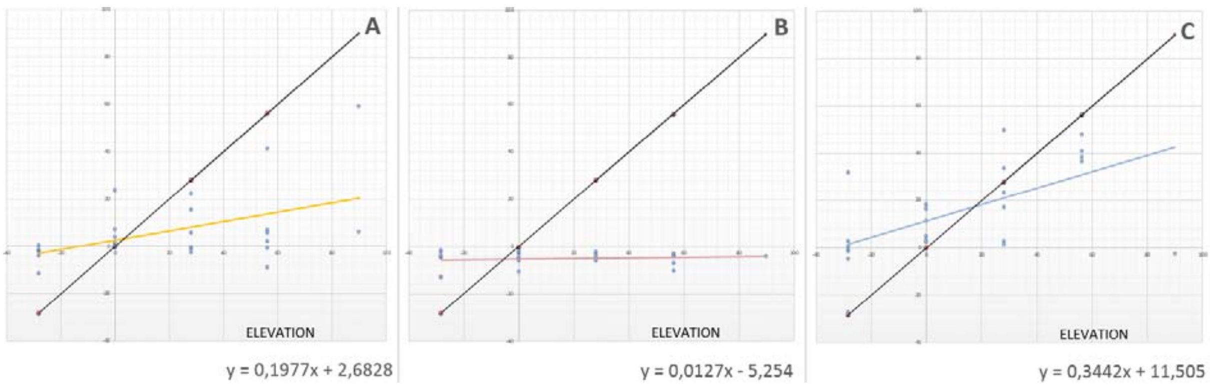



**Figure 5.12:** *Linear regression for elevation data for subject 012 for each HRTF condition*

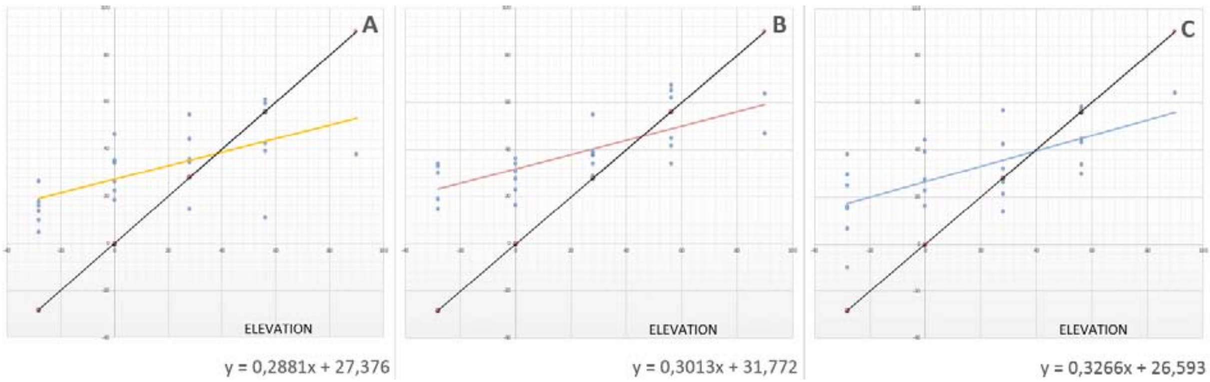



**Figure 5.13:** *Linear regression for elevation data for subject 013 for each HRTF condition*

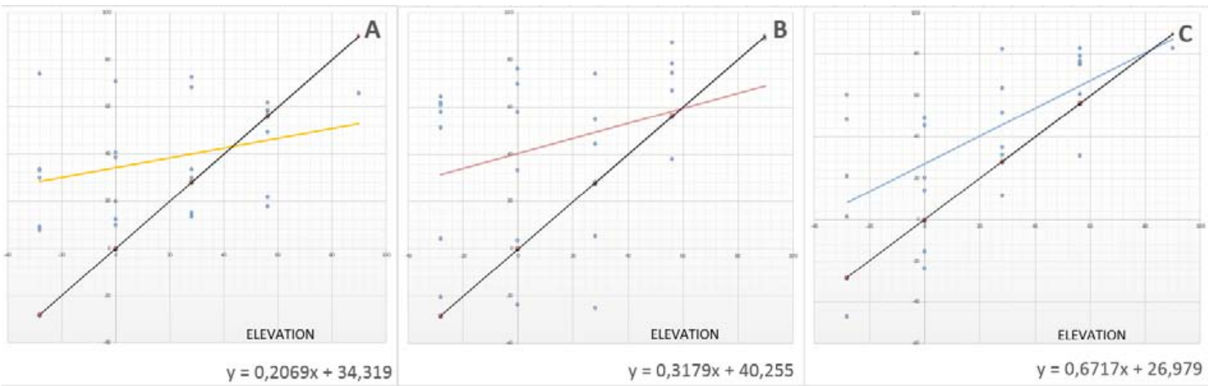



**Figure 5.14:** *Linear regression for elevation data for subject 014 for each HRTF condition*

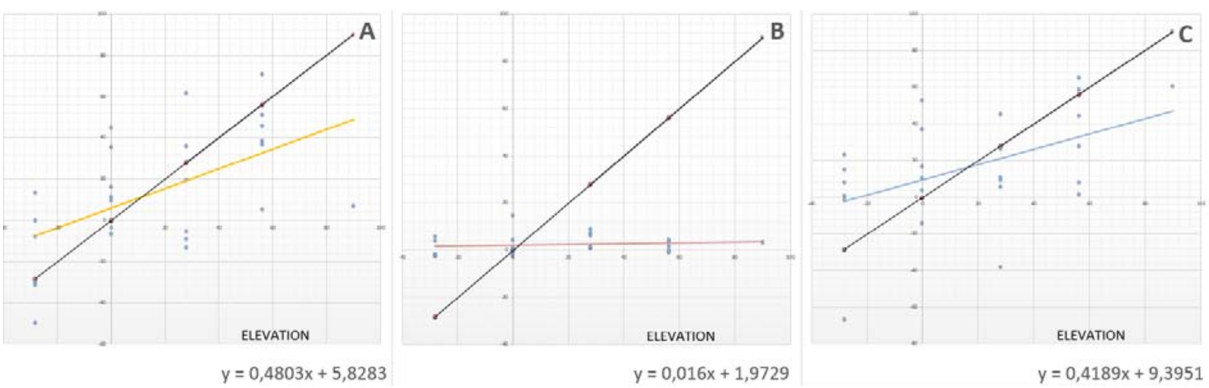



**Figure 5.15:** *Linear regression for elevation data for subject 015 for each HRTF condition*

| Slopes (NO 3-5-10) | | |
|---|---|---|
| | *A* | *B* |
| AVG | 0,260417 | 0,228333 |
| Variance | 0,017226 | 0,031132 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,014868 | |
| Stat t | 0,509036 | |
| P(T<=t) one tail | 0,310391 | |

**Table 5.4:** *Linear regression slopes. Paired T-Test on A-B conditions.*

| Slopes (NO 3-5-10) | | |
|---|---|---|
| | *A* | *C* |
| AVG | 0,260417 | 0,454583 |
| Variance | 0,017226 | 0,033848 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,537055 | |
| Stat t | -4,24232 | |
| P(T<=t) one tail | 0,000692 | |

**Table 5.5:** *Linear regression slopes. Paired T-Test on A-C conditions.*

| Slopes (NO 3-5-10) | | |
|---|---|---|
| | *B* | *C* |
| AVG | 0,22833333 | 0,454583 |
| Variance | 0,03113206 | 0,033848 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,57151238 | |
| Stat t | -4,6942699 | |
| P(T<=t) one tail | 0,0003281 | |

**Table 5.6:** *Linear regression slopes. Paired T-Test on B-C conditions.*

| FrontBack (NO 3-5-10) | | |
|---|---|---|
| | *A* | *B* |
| AVG | 2,27 | 3,93 |
| Variance | 2,78 | 28,50 |
| Num Occurrences | 15 | 15 |
| Pearson Correlation | 0,010164 | |
| Stat t | -1,15757 | |
| P(T<=t) one tail | 0,133204 | |

**Table 5.7:** *Front-back confusion errors. Paired T-Test on A-B conditions.*

| FrontBack (NO 3-5-10) | | |
|---|---|---|
| | *A* | *C* |
| AVG | 2,00 | 1,58 |
| Variance | 2,00 | 0,99 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,129055 | |
| Stat t | 0,890229 | |
| P(T<=t) one tail | 0,196199 | |

**Table 5.8:** *Front-back confusion errors. Paired T-Test on A-C conditions.*

| FrontBack (NO 3-5-10) | | |
|---|---|---|
| | *B* | *C* |
| AVG | 2,08 | 1,58 |
| Variance | 1,90 | 0,99 |
| Num Occurrences | 12 | 12 |
| Pearson Correlation | 0,02757386 | |
| Stat t | 1,03175391 | |
| P(T<=t) one tail | 0,16216962 | |

**Table 5.9:** *Front-back confusion errors. Paired T-Test on B-C conditions.*

# Bibliography

[1] V Ralph Algazi, Richard O Duda, Dennis M Thompson, and Carlos Avendano. The cipic hrtf database. In *Applications of Signal Processing to Audio and Acoustics, 2001 IEEE Workshop on the*, pages 99–102. IEEE, 2001.

[2] A. Andreopoulou, D. Begault, and B. Katz. Inter-Laboratory Round Robin HRTF Measurement Comparison. *IEEE Journal of Selected Topics in Signal Processing*, PP(99):1–1, 2015.

[3] Alberto Bedin. Head related transfer function selection techniques applied to multimodal environments for spatial cognition. 2013.

[4] Karsten Bormann. Presence and the utility of audio spatialization. *Presence: Teleoperators and Virtual Environments*, 14(3):278–297, 2005.

[5] Yashar Deldjoo and Reza Ebrahimi Atani. A low-cost infrared-optical head tracking solution for virtual 3d audio environment using the nintendo wii-remote. *Entertainment Computing*, 12:9–27, 2016.

[6] Avanzini Federico. Sound in space. 2009.

[7] Michele Geronazzo. *Mixed structural models for 3D audio in virtual environments*. Ph.D. Thesis, University of Padova, Padova, Italy, April 2014.

[8] Michele Geronazzo, Jari Kleimola, and Piotr Majdak. Personalization support for binaural headphone reproduction in web browsers. In *Proc. 1st Web Audio Conference*, 2015.

[9] Michele Geronazzo, Enrico Peruch, Fabio Prandoni, and Federico Avanzini. Improving elevation perception with a tool for image-guided head-related transfer function selection. In *(DAFx-17) proceedings of the 20th International Conference on Digital Audio Effects, Sept. 5-9, September, UK*, 2017. Submitted for publication.

[10] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Estimation and modeling of pinna-related transfer functions. In *Proc. of the 13th Int. Conference on Digital Audio Effects (DAFx-10)*, pages 6–10, 2010.

[11] Michele Geronazzo, Simone Spagnol, and Federico Avanzini. Mixed Structural Modeling of Head-Related Transfer Functions for Customized Binaural Audio Delivery. In *Proc. 18th Int. Conf. Digital Signal Process. (DSP 2013)*, pages 1–8, Santorini, Greece, July 2013.

[12] Michele Geronazzo, Simone Spagnol, Alberto Bedin, and Federico Avanzini. Enhancing vertical localization with image-guided selection of non-individual head-related transfer functions. In *Acoustics, Speech and Signal Processing (ICASSP), 2014 IEEE International Conference on*, pages 4463–4467. IEEE, 2014.

[13] Akio Honda, Hiroshi Shibata, Jiro Gyoba, Kouji Saitou, Yukio Iwaya, and Yôiti Suzuki. Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game. *Applied Acoustics*, 68(8):885–896, 2007.

[14] Yukio Iwaya. Individualization of head-related transfer functions with tournament-style listening test: Listening with others ears. *Acoustical science and technology*, 27(6):340–343, 2006.

[15] Brian FG Katz and Gaëtan Parseihian. Perceptually based head-related transfer function database optimization. *The Journal of the Acoustical Society of America*, 131(2):EL99–EL105, 2012.

[16] Alexander Lindau, Vera Erbes, Steffen Lepa, Hans-Joachim Maempel, Fabian Brinkman, and Stefan Weinzierl. A spatial audio quality inventory (saqi). *Acta Acustica united with Acustica*, 100(5):984–994, 2014.

[17] Piotr Majdak, Peter Balazs, and Bernhard Laback. Multiple Exponential Sweep Method for Fast Measurement of Head-Related Transfer Functions. *JAES*, 55(7/8):623–637, July 2007.

[18] Erik Murphy-Chutorian and Mohan Manubhai Trivedi. Head pose estimation in computer vision: A survey. *IEEE transactions on pattern analysis and machine intelligence*, 31(4):607–626, 2009.

[19] Vikas C Raykar, Ramani Duraiswami, and B Yegnanarayana. Extracting the frequencies of the pinna spectral notches in measured head related impulse responses. *The Journal of the Acoustical Society of America*, 118(1):364–374, 2005.

[20] Ludivine Sarlat, Olivier Warusfel, and Isabelle Viaud-Delmon. Ventriloquism aftereffects occur in the rear hemisphere. *Neuroscience letters*, 404(3):324–329, 2006.

[21] Patrick Satarzadeh. *A study of physical and circuit models of the human pinnae*. PhD thesis, UNIVERSITY OF CALIFORNIA DAVIS, 2006.

[22] Simone Spagnol, Michele Geronazzo, and Federico Avanzini. On the relation between pinna reflection patterns and head-related transfer function features. *IEEE transactions on audio, speech, and language processing*, 21(3):508–519, 2013.

[23] Elizabeth M Wenzel, Marianne Arruda, Doris J Kistler, and Frederic L Wightman. Localization using nonindividualized head-related transfer functions. *The Journal of the Acoustical Society of America*, 94(1):111–123, 1993.

[24] Satoshi Yairi, Yukio Iwaya, and Yôiti Suzuki. Individualization feature of head-related transfer functions based on subjective evaluation. In *Proc. of International Conference on Auditory Display (ICAD2008), Paris*, 2008.

[25] DYN Zotkin, Jane Hwang, R Duraiswaini, and Larry S Davis. Hrtf personalization using anthropometric measurements. In *Applications of Signal Processing to Audio and Acoustics, 2003 IEEE Workshop on.*, pages 157–160. Ieee, 2003.