HELSINKI UNIVERSITY OF TECHNOLOGY
Faculty of Electronics, Communications and Automation
Department of Signal Processing and Acoustics

Ville Riku Johannes Riikonen

User-related Acoustics in a Two-way Augmented Reality Audio System

Master's Thesis submitted in partial fulfillment of the requirements for the degree of Master of Science in Technology.

Espoo, April 24, 2008

Supervisor: Professor Matti Karjalainen
Instructor: Miikka Tikander, M.Sc.

HELSINKI UNIVERSITY OF TECHNOLOGY		ABSTRACT OF THE MASTER'S THESIS
Author:	Ville Riku Johannes Riikonen	
Name of the thesis:	User-related Acoustics in a Two-way Augmented Reality Audio System	
Date:	April 24, 2008	**Number of pages:** 77
Faculty:	Faculty of Electronics, Communications and Automation	
Professorship:	S-89	
Supervisor:	Prof. Matti Karjalainen	
Instructor:	Miikka Tikander, M.Sc.	

In Augmented Reality Audio (abbreviated ARA), the real sound environment of the user is extended with virtual sound images. The surrounding environment is produced with a special ARA headset. The ARA headset consists of miniature microphones, which are integrated to earphone elements in both ears. The signals from the microphones should go unaltered directly to the earphones. This way the user should be able to hear an exact copy of the surrounding sound environment. This copy is called the *pseudoacoustic* environment. However, the headset causes coloration to the pseudoacoustics.

After the pseudoacoustic sound field has been generated, new virtual sound objects can be added to the user's sound environment. This is done with the help of a special device, the ARA mixer. The synthetic sound objects can be almost anything (speech from a remote user, conversation of a remote group, music, alarm signals, calendar marks, etc.) and they can be heard separately from the real sound environment or the user might not even distinguish between the real and synthetic sounds. To be able to control the virtual sound events, the system has to keep track of the user's movement and rotation. This is called *head-tracking and positioning*.

The main issues discussed in this thesis are the sound quality of the pseudoacoustic representation, the rendering of the virtual sound images in a controlled manner, and ARA platform development. A novel ARA mixer prototype is constructed and evaluated. It has equalization properties to enhance the pseudoacoustic sound quality. Also acoustic head-tracking and positioning techniques are introduced. A platform for ARA application testing with a two-way communication system and a possibility for head-tracking is also introduced.

Keywords: Augmented Reality Audio (ARA), ARA headset, ARA technology, pseudoacoustics, head-tracking, positioning

TEKNILLINEN KORKEAKOULU	DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Ville Riku Johannes		
Työn nimi:	Käyttäjäkeskeinen akustiikka kahdensuuntaisessa Lisätyn Audiotodellisuuden järjestelmässä		
Päivämäärä:	24.4.2008	**Sivuja:**	77
Tiedekunta:	Elektroniikan, tietoliikenteen ja automaation tiedekunta		
Professuuri:	S-89		
Työn valvoja:	Prof. Matti Karjalainen		
Työn ohjaaja:	DI Miikka Tikander		

Lisätyn audiotodellisuuden (LAT) sovelluksissa käyttäjän ympäröivää todellista äänikenttää laajennetaan keinotekoisilla ääniobjekteilla. Ympäröivä äänikenttä luodaan erityisen LAT-kuulokkeen avulla. LAT-kuuloke koostuu miniatyyrimikrofoneista, jotka on liitetty normaaleihin kuulokkeisiin molemmissa korvissa. Mikrofonisignaalin pitäisi mennä muuttumattomana suoraan kuulokkeeseen. Täten käyttäjän pitäisi kuulla tarkka kopio ympäröivästä äänikentästä. Kopiota kutsutaan *pseudoakustiseksi* ympäristöksi. LAT-kuuloke kuitenkin värittää pseudoakustista ympäristöä.

Kun pseudoakustinen äänikenttä on luotu, voidaan uusia keinotekoisia ääniobjekteja lisätä käyttäjän ympäristöön. Tämä tehdään siihen suunnitellun laitteen, LAT-mikserin, avulla. Keinotekoiset ääniobjektit voivat olla oikeastaan mitä vain (toisaalla olevan ihmisen puhetta, toisaalla olevan ryhmän keskustelua, musiikkia, hälytysääniä, kalenterimerkintöjä, jne.), ja ne voidaan kuulla erillisinä oikeasta äänikentästä tai käyttäjä ei välttämättä edes erota oikeita ääniä keinotekoisista. Jotta keinotekoisia ääniä voitaisiin hallita, täytyy systeemin tietää käyttäjän liikkeistä ja rotaatiosta. Tätä kutsutaan *päänliikkeiden jäljittämiseksi ja paikannukseksi*.

Tässä diplomityössä käsitellään pseudoakustiikan äänenlaadun parantamista, keinotekoisten äänilähteiden hallintaa sekä LAT-alustan kehittämistä. Uudenlainen LAT-mikseriprototyyppi suunnitellaan, rakennetaan ja arvioidaan. Siinä on ekvalisointimahdollisuus pseudoakustisen äänenlaadun parantamiseksi. Myös akustista päänjäljittämistä ja paikannusta käsitellään. Alusta LAT-sovelluksien testaamiseen esitellään. Siinä on kahdensuuntaiseen kommunikointiin tarvitta ohjelma sekä mahdollisuus päänliikkeiden seurantaan.

Avainsanat: Lisätty Audiotodellisuus, kuulokkeet, lisätyn todellisuuden teknologia, pseudoakustiikka, päänliikkeiden jäljittäminen, paikannus

Acknowledgements

This Master's thesis has been done in the Department of Signal Processing and Acoustics in cooperation with the Department of Media Technology and with funding from Nokia Research Center.

I want to thank Professor Matti Karjalainen for excellent guidance and supervision. Matti's comments and instructions have been vital and help has been always available when needed.

I wish to thank M.Sc. Miikka Tikander for useful assistance and practical guidance. Miikka has brought his knowledge and experience available to the project.

I would also like to thank my co-workers Tapio Lokki, Lauri Savioja, Sampo Vesa, and Laura Seppänen at the Department of Media Technology for cooperation and shared efforts.

I also appreciate all the help and support from the staff and co-researchers of the Department of Signal Processing and Acoustics. Your enthusiasm and help have been more than welcome.

My gratitude also goes to Matti Hämäläinen, Riitta Väänänen, Julia Turku, and Mika Ristimäki at Nokia Research Center for bringing fresh ideas and insights to the project.

Finally, I would like to thank my dear brother Janne Riikonen for support and understanding, Mikko Honkala for "scientific inspiration", Antti "Ukki" Riikonen for participating in usability evaluation, and my parents Ritva Riikonen and Jukka Riikonen for all kinds of support during my studies. Without You, this would have not been possible.

Otaniemi, April 24, 2008

Ville Riku Johannes Riikonen

Contents

Abbreviations

ARA	Augmented Reality Audio
MARA	Mobile Augmented Reality Audio
KAMARA	Killer Applications in Mobile Augmented Reality Audio
HRTF	Head-Related Transfer Function
ITD	Interaural Time Difference
ILD	Interaural Level Difference
6-DOF	Six Degrees of Freedom
LED	Light Emitting Diode
GPS	Global Positioning System
WLAN	Wireless Local Area Network
MP3	Mpeg 1 Audio Layer 3
AR	Augmented Reality
VR	Virtual Reality
SPL	Sound Pressure Level

Chapter 1

Introduction

This thesis is closely related to a project called KAMARA+ (Killer Applications of Mobile Augmented Reality Audio). The first stage of the project was carried out from December 2001 to December 2004 in five phases and it was a shared effort between Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing and Telecommunications and Software and Multimedia Laboratory (TML), and the Nokia Research Center (NRC).

Since then, the project has been ongoing both at Nokia Research Center and Helsinki University of Technology. From the beginning of 2005, two doctoral students have participated in KAMARA project with academic funding. The collaboration between NRC and the two laboratories at Helsinki University of Technology continued from autumn 2007, including this thesis and another thesis done at TML.

This thesis, along with the other one made at the TML, includes development of a two-way communication system. However, there are differences in the aspects that are emphasized. At TML the emphasis is more on software architecture and video based head-tracking whereas this thesis concentrates on the user-related acoustics, especially on the sound quality issues of Augmented Reality Audio headsets and acoustic head-tracking.

The reason and motivation for this type of platform approach is that the project has broadened so much during past years that there is a need for a common platform, where different types of experiments and tests can be reliably and consistently executed and reported. Earlier, the experiments have been done quite isolated from each other at various sites.

In augmented reality audio, the real sound environment of the user is extended with virtual sound images. The fact that the user is still capable of hearing the natural sound environment makes the fundamental difference between augmented reality (AR) and pure virtual reality (VR).

The main idea behind the KAMARA-project was first documented in article [9]. What makes the concept very interesting and innovative, is that it mixes virtual reality and acoustic reality in a way which could have a wide variety of different applications. These applications include both entertainment and utility applications.

The main goal of this thesis was to study user-related acoustics in ARA applications and build and test a two-way communication platform between the Department of Media Technology (former Telecommunications and Software and Multimedia Laboratory) and the Department of Signal Processing and Acoustics (former Laboratory of Acoustics and Audio Signal Processing). The focus in the user-related acoustics was in the headset and ARA mixer technology. New headsets and mixers were constructed and evaluated. The emphasis in the platform development was in getting the communication to work between two workstations and to construct and evaluate a head-tracking system.

Chapter 2 of this thesis concentrates on introducing the ARA technology and applications in general. In Chapter 3, the development and construction of new ARA headsets and mixers is documented. Chapter 4 presents principles behind acoustic head-tracking and positioning. Also platform-related issues are covered. In Chapter 5, the outcomes of the work are tested and analyzed and Chapter 6 presents the conclusions of this study and suggests what still needs to be done.

Chapter 2

Overview of the ARA Technology

2.1 The Concept of Augmented Reality Audio

In Augmented Reality Audio, the real sound environment of the user is extended with virtual sound images [6], so it is not pure virtual reality. To be able to embed virtual objects to the natural sound environment, special equipment and some signal processing is needed. In [9] the concept of Augmented Reality Audio (ARA) headsets has been introduced.

The ARA headset consists of miniature microphones, which are integrated to earphone elements in both ears. The signals from the microphones should go unaltered and with minimal latency directly to the earphones. This way the user should be able to hear a copy of the surrounding sound environment. This copy is called the pseudoacoustic environment [9]. Of course this is the ideal case; in reality the headset system is not completely transparent and causes some coloration to the pseudoacoustic representation of the real environment. The same representation can also be transmitted to other users in another environment and they can hear the same pseudoacoustic sound field as the person from which the sounds are originally recorded.

After the pseudoacoustic sound field has been generated, new virtual sound objects can be added to the user‘s sound environment. This is done with the help of a special device, the ARA mixer (see Chapter 3.5). The synthetic sound objects can be almost anything (speech from a remote user, conversation of a remote group, music, alarm signals, calendar marks, etc.) and they can be heard separately from the real sound environment or at the other extreme the user might not even know, which of the sounds are real and which synthetic [9].

Figure 2.1 shows the basic system diagram for an augmented reality audio application [9]. If the new sound objects are to be heard completely embedded in the real sound environment, there is a need to keep track of the person’s movement and position. There are many ways to do this and the process is called head-tracking and positioning [34] (see Chapter 2.4).

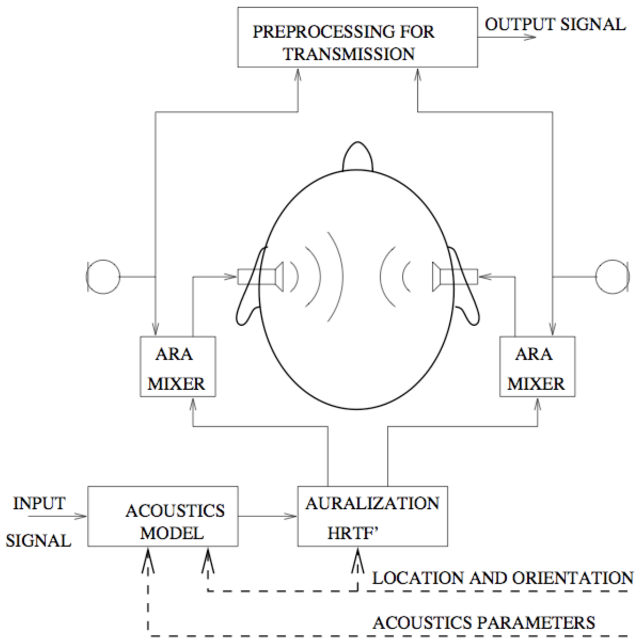

Figure 2.1: System diagram for an augmented reality audio application (after [9]).

2.2 Hardware of ARA Technology

In its most stripped-down version, the hardware needed for ARA applications is quite simple. In order to hear the natural sound environment, a special headset with integrated binaural microphones is needed. The microphone signals are routed directly to the earphones through a mixer with microphone preamplifiers and headphone amplifiers. Also equalization can be done in the mixer (see Section 3.5). Now it is possible to create the pseudoacoustic representation of the real sound environment. It is also possible to add virtual sound objects to the pseudoacoustic environment by simply mixing the new audio signals to the pseudoacoustic representation with the help of the ARA mixer. The added signals can be obtained from a computer, a PDA, or for example a mobile phone. It is also possible to send the binaural microphone signals from the mixer to a computer and finally to a distant user. In a more advanced system, where sound objects are to be placed in specific coordinates in the natural sound field or moved around the user, head and motion tracking is needed (see Section 2.4 and Chapter 4).

2.2.1 ARA Headset

The Augmented Reality Audio headset is quite similar to a hearing aid. It has miniature microphones to capture the sound field and transducers to reproduce the microphone signals. The quality of the microphones and transducers is critical in terms of creating a vivid pseudoacoustic representation. The headset should also be comfortable enough to be worn by the user for longer periods of time.

Since there are no commercial ARA headsets available, they have to be constructed. One way is to add small electret microphones to commercial headphones (usually in-ear type of headphones, for different types of headphones see Section 3.1.2). This has also been the approach in prior research (for example in [35, 9]). The earphones used in these studies have mostly been loosely-fit in-ear headphones. This type of ARA headset is shown in Figure 2.2.

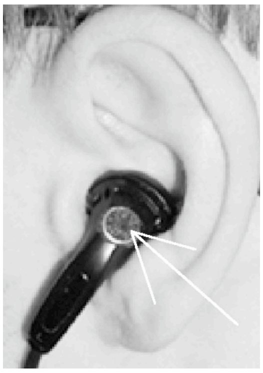

Figure 2.2: An experimental ARA headset. The arrow points at the microphone (after [35]).

In this study we primarily use more novel insert-type of headphones, which are based on active noise canceling headphones. The advantage is that they already have integrated microphones. The only construction that has to be made is to disconnect the noise canceling control box and make new wiring to route microphone signals to the ARA mixer. The headphone used in this study is Philips SHN2500, which is shown in Figure 3.11. For more details about the headphone, see Section 3.4.1.

2.2.2 ARA Mixer

The Augmented Reality Audio mixer is an essential part of the ARA hardware. The mixer takes care of the signal routing and mixing and acts like an interface between the user and the outside world. The mixer has the following main tasks:

- Routing the microphone signals to the earphones
- Routing the microphone signals to an external system like a computer or a PDA
- Receiving incoming audio signals from an external system and adding them to the pseudoacoustic representation

There can also be more properties (the first two can be found in the novel ARA mixer described in Section 3.5):

- Analog equalization possibilities
- Connectors for using an external equalization device
- Phase shift (noise canceling)

It is desirable that the ARA mixer is so small and light-weight that the user can keep it in a pocket all the time. The power consumption should also be as low as possible to ensure a long operating time. Wireless connection for communication with the external devices would also be beneficial and enhance the usability. For more details about the ARA mixer and the development made in this study, see Section 3.5.

2.2.3 Head-tracker

A sophisticated ARA system requires some sort of a head and motion tracker. The reason for this is that without a head-tracking system, the virtual sound objects or sound field would move along with the head. In order to keep the virtual sound source stationary or to move it in the sound-field in a controlled manner, it is necessary to keep track of the motion of the user. There are many ways to do head-tracking, and there are some commercial head-tracking devices available, but they lack in robustness and operating range. In this thesis, acoustic head-tracking based on stationary anchor sources is studied. For more details about head-tracking, see Section 2.4 and Chapter 4.

2.2.4 Computer or a PDA

To be able to add virtual sound objects to the pseudoacoustic representation of the natural sound environment, some kind of signal source is needed. This source sends the virtual sound signals to the ARA mixer, which adds them to the binaural signals. The sound source can be a computer, a PDA, or for example a mobile phone. If head and motion tracking is used, it can also be conducted in the same device. The signal processor also receives the binaural recordings from the ARA mixer and can route them to a distant user. If there are many users in the same space, the signal processor takes care of routing and allocation of different signals.

2.3 Basic Concepts of Binaural Hearing

Humans normally have two ears, which helps in interpreting information from the surrounding acoustic environment. This is called binaural hearing. Gathering information would be much harder with only one ear (monaural hearing). With the help of two ears, it is possible to obtain information about sound sources based on a complex series of direct sounds, early reflections, and reverberation of the space. Humans can separate sound sources and also estimate their direction and distance [13]. This kind of spatial and directional hearing is an important part of ARA applications, since binaural cues recorded from one user can be utilized by other users. Also head-tracking is based on binaural cues (see Chapter 2.4). If these cues are missing, localization is difficult and often the sound source is located inside the listener's head. This is called lateralization and happens for example in headphone listening of commercial material, which is originally intended to be used with loudspeakers. Lateralization can be avoided by manipulating the signals before playing them back with headphones or by playing back signals recorded already binaurally like in many ARA applications [2].

What is also worth noticing, especially considering ARA applications, is that binaural cues already exist at the ear canal entrance where the microphones of the ARA headset usually are. That is to say that the ear canal itself does not add any extra information needed in localization. The ear canal resonance is always the same, no matter what the angle of the arriving sound signal is [13].

2.3.1 Interaural Time Difference - ITD

One important binaural cue is the Interaural Time Difference (abbreviated ITD). ITD can be easily explained by looking at Figure 2.3. When the sound source is in the center, the wavefront has the same distance to both ears. If the source is located off-axis, the wavefront has a shorter distance to the right ear. Now the excitation signal has to travel a shorter path to reach the right ear, which takes less time than to reach the left ear. With the help of this interaural time difference, humans can estimate the direction of the sound source [2]. If the average head size is assumed to be 18 cm, the maximum time difference is approximately

700 μs, which occurs when the sound comes directly from the side of the head. ITD also depends on the vertical angle.

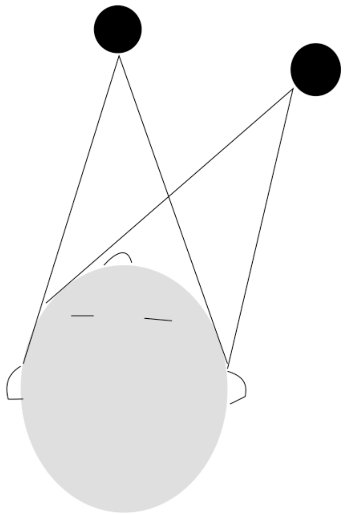

Figure 2.3: A listener with a sound source directly ahead and displaced to right.

2.3.2 Interaural Level Difference - ILD

It is clearly evolutionary that humans have ears located on the opposite sides of the head. This kind of positioning maximizes the interaural differences of the sound events occurring around the listener. Because ears are located on opposite sides of the head, the wavefront is different at the two ears. One of the most important differences is the Interaural Level Difference (abbreviated ILD), which is a frequency-dependent cue [2].

ILD, as the name implies, measures the sound pressure level difference at the two ears caused by shadowing of the head to the wavefront. The sound pressure level is higher on the side of the head where the sound comes from [13]. ILD is also frequency dependent. At low frequencies, where the wavelength is much longer than the dimensions of the head, the shadowing effect is almost absent and localization based on interaural level differences is difficult. The head does not cause an obstacle to the wavefront, instead the sound wave diffracts and bends around the surface of the head.

At mid-range and especially at high frequencies where the dimensions of the head are much larger than the wavelength, the shadowing effect of the head is stronger and thus the level differences are more noticeable. Based on an average sized head, the corner frequency for ILD to become dominant is approximately 1.5 kHz and the shadowing effect increases with increasing frequency [2]. The effect is shown in Figure 2.3. It can be noticed that the sound source, which is directly ahead the listener causes similar signals to both ears. The other sound source is located more on the right side and thus the signal reaching the left ear is interfered by the head while there are no obstacles on the signal route to the right ear [2].

2.3.3 Head-Related Transfer Functions

ITD or ILD alone do not explain the whole process of localization or spatial and directional hearing. The needed extra cues come from the human anatomy. Head, outer ear, shoulders and upper torso also affect the sound signal entering the ear canal. The signal is altered according to the incoming angle by reflections, coloration and shadowing caused by the human body. This individual effect can be depicted by the so called Head-Related Transfer Functions (abbreviated HRTF) [2].

To be able to describe the direction of the incoming wavefront accurately, the space around the listener's head has to be divided into sections. The origin of the coordinate system is inside the listener's head between the ears. The median plane divides the head in two sections vertically and the horizontal plane divides it horizontally. The frontal plane goes though the ears vertically. The azimuth and elevation angles determine the direction of the sound source and distance (r) to the source has to be also known in order to unambiguously specify the location in the spherical polar coordinate system [13].

HRTF is the transfer function of the signal route from a point source to the ear canal entrance or inside the ear canal measured in free-field. The transfer function is individual for each subject and can be measured in an anechoic chamber with source loudspeakers around the listener and miniature microphones installed either at the ear canal entrance or inside the ear canal. If the microphone is at the ear canal entrance, the ear canal can be either blocked (as in most ARA headsets) or open, which has to be taken into account.

For the best result, individual HRTFs should be used, but for practical reasons averaged HRTFs from a set of listeners is often used. This saves time and money. Another option is to use HRTFs recorded with a dummy head and microphones inside it's ears. A dummy head is a head and torso simulator, which is based on average adult anthropometric data [2]. Figure 2.4 shows the Brüel&Kjær HATS 4128 head and torso simulator.

When filtered with either individual or averaged HRTF filters, one can create spatial and directional effects to normal mono or stereo recordings. This is often used in augmented and virtual reality audio. In ARA applications, binaurally recorded sound events are transmitted to other users. Of course these recordings have the individual HRTF characteristics of the user from which they are recorded, and how well these recordings work with other people depends on the listener's own HRTF functions.

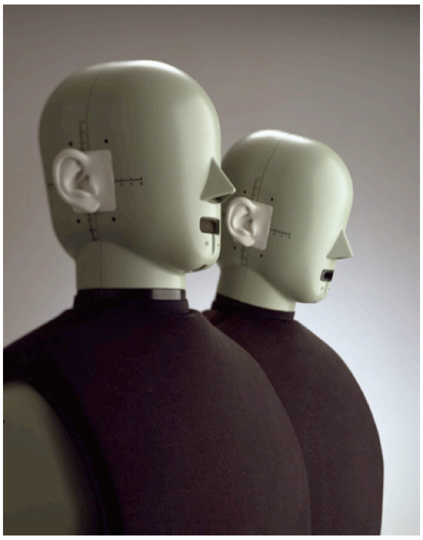

Figure 2.4: The Brüel&Kjær HATS 4128 head and torso simulator (after [25]).

2.3.4 Localization

To be able to localize a sound object, humans need the information about direction and distance. When only one source is considered and the listening happens in free-field, only a few parameters is enough to produce localization. These are horizontal and elevation angles and distance. Elevation is less accurate to evaluate than horizontal position. The estimation of distance is also difficult and is primarily based on sound pressure, not to the real distance (at least in free-field conditions). In normal listening situations, the reverberation and early reflections help in estimating the size of the room and distance to sound objects [13].

2.4 Positioning and Head-tracking

In most virtual and augmented reality audio applications, it is vital to know the exact position of the user and also the orientation of the head (where the user is looking at). This enables total control of the direction and distance of sound events and control of ambience [13]. For example, if the user is in a room with an ARA headset and there is the ARA system needed for importing virtual sound events to the user's ears, it is necessary to keep track of the user's motion and rotation. Otherwise the virtual sound objects or the sound field would move along as the head is turning and moving. In order to keep the virtual sound source stationary or to move it in the sound field in a controlled way, it is necessary to keep track of the user's motion. In Figure 2.5, coordinates for pitch, yaw, and roll along with the position information x, y, and z are shown.

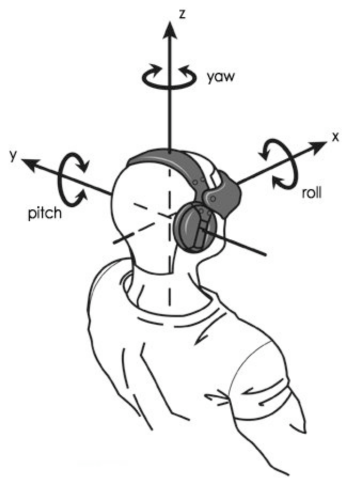

Figure 2.5: Coordinates for pitch, yaw, roll, x, y, and z in a head-tracking system (after [2]).

There are many ways to do motion- and head-tracking. Common for all tracking systems is the need for a source, a sensor, and a hardware unit, which is connected to a computer. Some of the most popular tracker technologies are based on electromagnetic sensors. One of these commercial applications is the Polhemus Fastrak. Other systems often use optical, mechanical, or acoustic tracking. Different systems have their advantages and disadvantages considering the area where it can function, accuracy, and response latency [2].

2.4.1 Head-tracking Based on Electromagnetic Sensors

The most common head-tracking technology is based on electromagnetic sensors. The system is usually constructed so that a sensor is attached to the user's body. Often it is on the top of the head of the user. The sensor measures the low-frequency magnetic field, which is generated by a static source somewhere in the room. The sensor is attached to hardware, which can calculate the relative position and direction of the sensor in relation to the source, who's coordinates are known. This enables the virtual sound objects to remain fixed or move in a controlled manner independent of head movement [2].

The biggest problem with this kind technology is the interference from other devices that can produce magnetic fields in the room. Also metal objects, like office furniture, can disrupt magnetic fields [1].

2.4.2 Acoustic Head-tracking Based on Binaural Signals

The User As a Sound Source

When the user acts as a sound source for the tracking system, the configuration is often called an *outside in* system. In this kind of system, there are static microphone arrays installed in the room and the user is sending some known signals. The signals are captured by the microphone arrays, and based on the arrival times and level differences, the tracking system calculates and estimates the position and orientation of the target [21].

There are downsides to this system. For example, it includes special hardware required by the use of ultrasonic frequencies, which is common in this kind of systems. Another problem is that the positioning information is processed at the receiver, not at the user (target), where the information is more often needed. The user has to carry loudspeakers, which have a higher power consumption than the miniature microphones, which are used in the system presented in the next Chapter [34].

Stationary Anchor Sources

With stationary anchor sources, the system is otherwise similar to the *outside in* system, but reverse in terms of the source and the receiver of the reference signal. When stationary anchors (usually loudspeakers) are used and the target is wearing the receiving microphones, the system is called an *inside out* system [34]. In this kind of setup, the anchors in the environment can be either known or unknown. The inside out system is a very natural way of tracking the target's motion in an ARA environment since the user is already wearing miniature microphones in the ears, so no additional hardware is needed on the user side.

The complexity of the system is lower if known anchor sources, like loudspeakers, are used. In many places there are already loudspeakers, which can be utilized. Loudspeaker setups exist in most classes, auditoriums, offices, cars, living rooms, and home offices. More robustness could be obtained by utilizing also other static sound sources like air-conditioning outlets, the fan of a computer or a printer.

2.4.3 Optical Head-tracking

Head-tracking based on optical devices is often constructed in either of the following ways. In one construction, one or more cameras are placed on top of the head motion detector. There is also a set of infrared LEDs (Light Emitting Diode) above the user at fixed locations. The LEDs send known signals (often pulses), which reflect from the head motion detector. The camera captures these reflections and based on this information, the exact position and orientation of the head can be calculated in a computer.

In another kind of setup, there are also cameras on the ceiling, but the LEDs are placed on the head motion tracker and move as the user moves. An advantage in both setups is the minimal latency. The downside is that the line of sight between the cameras and the LEDs can be obscured and interfere with the tracking process [1].

2.4.4 Mechanical Head-tracking

One possible tracking technique is the mechanical head-tracking. In this kind of system, the tracking is based on a physical connection between the target and the known reference point. They are very accurate and the update rate is high. The most severe disadvantage is the short operating distance and lack of mobility.

Often mechanical trackers consist of a light-weight arm, which is connected to a head-band through a control box. There are encoders in the arm that measure the changes in position and rotation with respect to the reference point.

2.4.5 Alternative Positioning and Head-tracking Systems

Positioning Based on GPS

As can be noticed from the previous sections, there are many different tracking systems, but most of them suffer from poor mobility and short operating range. One possibility would be positioning based on GPS (Global Positioning System). The advantage of this method is wide-range mobility. The downside is it's lack of ability to track the user's orientation (direction of facing). Therefore it is rather a positioning system than a head-tracker. There are also problems concerning precision and robustness. These are particularly severe problems inside buildings [14]. Because many of the ARA applications would be used indoors, GPS positioning is out of question at least for now.

Video-based Head-tracking and Positioning

Also video-based head-tracking could be one option for ARA purposes. It is in quite close relation to optical head-tracking presented in Chapter 2.4.3. The main idea in video tracking is to locate a moving object with the help of a video camera. The frames recorded by the camera are sent to a computer, which analyzes them based on a chosen algorithm and estimates the position and motion of the object. Motion prediction is also often used, which decreases the amount of computing [19].

Since object recognition from video frames has been problematic and the complexity of the vision-based approaches has been high, there have not been many commercial or robust systems available. However, there are a few more successful examples, but they are limited to a single specific task [19], for example a video-tracking system, which has video cameras in the ceiling and the user wears a certain colored hat. The object recognition program can find the hat from the video frames and thus estimate and predict the movement and orientation of the target.

Positioning Based on WLAN

One of the new and interesting possibilities for positioning could be the use of WLAN (Wireless Local Area Network). Since many modern mobile phones and other portable devices like MP3-players have built-in WLAN receivers and also many public places like airports, shopping centers, restaurants, and parks have WLAN access, it would be possible to track the movement of a target user with the help of this network. This could be beneficial for some of the ARA applications described in Chapter 2.5, but the disadvantage is the lack of ability to track the user's orientation (direction of facing) and poor resolution (at least at the time of this study).

2.5 Application Scenarios

The ARA headset and system allows many novel applications to be implemented. It offers a new way of human-to-human and human-to-machine communications. The applications can be either information services or communication services. An example of information service would be the Virtual Tourist Guide. This service would offer information and guidance to a tourist walking around in a city with an ARA headset on. A communication type of service would be for example Binaural Telephony or Virtual Meeting, where vivid telepresence would occur [20].

2.5.1 Binaural Telephony and Virtual Meeting

Virtual Meeting

In the globalized business world, it is very common that there is a lot of cooperation between people and teams all around the world. When face-to-face meetings are impossible to arrange, virtual meetings have to be held. One of the problems of video conferencing and traditional telephone discussions is the lack of telepresence and it can be very exhausting and difficult to follow a distant meeting this way. The ARA system allows a novel way of arranging virtual meetings.

Imagine a situation where one of the employees is somewhere outside the office and there is a meeting he/she would like to attend. Now, one of the attendants of the meeting wears an ARA headset and the teleworker receives the binaural signals recorded from the meeting.

This way the whole meeting room's auditory scene is transported to the teleworker, and he/she can easily separate and distinguish different attendants and who is talking at the time. This kind of application scenario has a number of different variations and can be extended and tailored to many situations. Figure 2.6 shows the basic idea in telepresence.

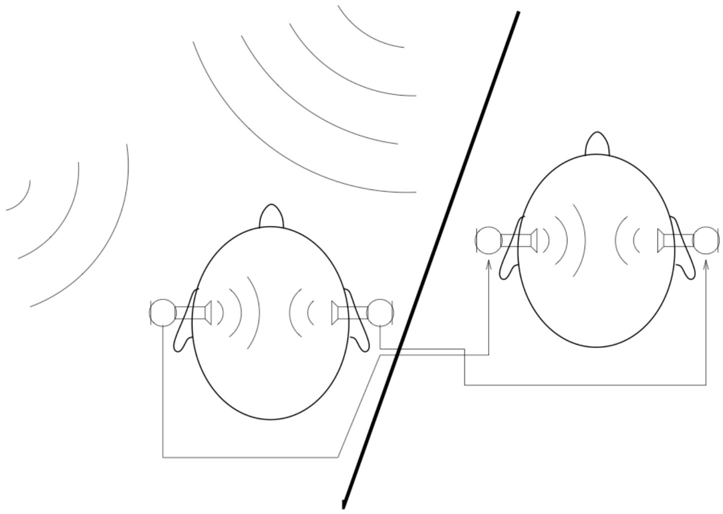

Figure 2.6: On the left: a user in a pseudoacoustic environment. On the right: a user experiencing telepresence of the far-end user (after [10]).

Binaural telephony

In binaural telephony, the binaural signals of both users are transmitted to each other. It is a voice-over-IP (VoIP) application, which offers a more realistic telepresence than traditional monophonic telephone. The whole soundscape of both users is transmitted to one another. The effect is the strongest when there are also other people and sound sources present in the room. The most obvious transport protocol for this kind of data is UDP [10]. Some compression algorithm could be used to save network capacity.

There are a few problems that concern both binaural telephony and virtual meetings. When the user wearing the ARA headset starts to talk, his/her voice is heard much louder by the far-end user than the voice of other people in the room (this is because other people are further away from the binaural microphones than the talker himself who is wearing the microphones). The voice is also located inside the far-end listener's head.

This can be avoided by using an algorithm, which detects when the wearer of the headset is talking and renders the sound so that it seems to come from outside the far-end listener's head and with a lowered level. The panning can be done with the help of HRTFs [10]. Figure 2.7 presents the rendering process. In the figure, the dashed wave fronts represent the virtual sound sources from another user and the solid lines are real sound sources. In the rendering process, the far-end user's voice is panned to the front of the near-end listener's head (the dashed wave front).

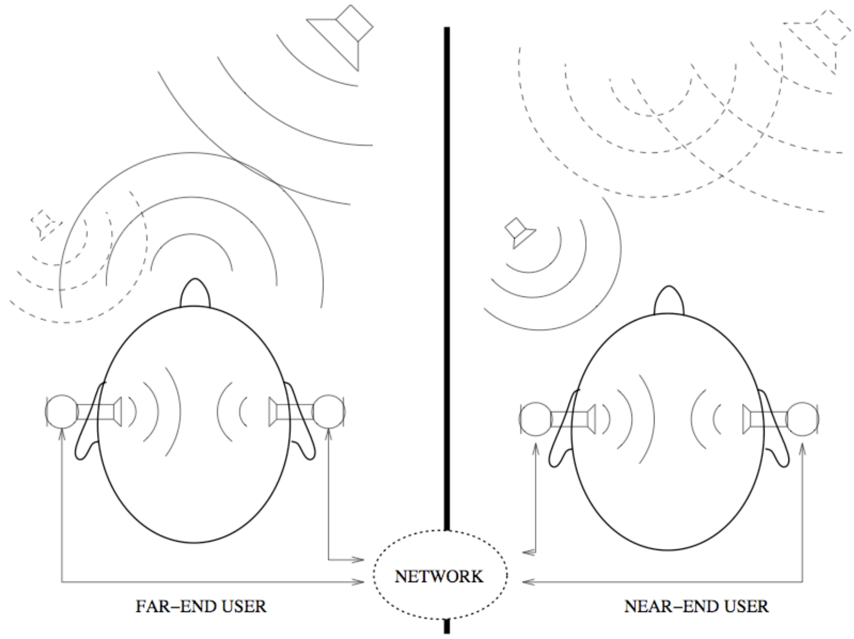

Figure 2.7: The rendering process of the far-end user's voice to the front of the near-end user (after [20])

Another problem with binaural telephony and virtual meetings is that if only one user in the meeting room is wearing the ARA headset, the others are not able to hear the far-end user's voice [10]. One option would be to use loudspeakers in the room to reproduce the far-end user's voice. Another option could be that all attendants in the meeting room could wear ARA headsets, but only one user's binaural signals would be sent to the far-end user. This way, all the people in the meeting room would also hear the far-end user's natural sound environment.

2.5.2 Acoustic Post-it Stickers

Another useful ARA application is the Acoustic Post-it Sticker [10]. The basic idea behind this application is the same as in the traditional yellow Post-it stickers. The user can leave messages to himself or to other people, and these messages can be location-based or object-based (attached to some object). For example, if a person leaves for holiday, he/she can leave an Acoustic Post-it sticker (a note) to the office door, which tells how long the person will be away. The announcement will be played through the ARA headset of the person who is standing at the door. This kind of auditory stickers would be very useful in workplaces. The Acoustic Post-it Sticker has been implemented in previous research, for example in [10, 20]. Positioning is required to keep track of user's motion.

2.5.3 Virtual Tourist Guide

Imagine walking around in a foreign city as a tourist. It would be very interesting to learn about different tourist attractions. However, tourist guides are expensive and can diminish the sense of freedom. Maps are practical, but they do not have much information and reading them is inconvenient. One interesting ARA application would be the Virtual Tourist Guide. It is a location-based application, which also needs positioning to keep track of the user's motion and perhaps head-tracking if the direction of the head is of importance (for example the Virtual Tourist Guide could tell something about the exact object the user is currently looking at). The idea is that the tourist could walk around the city with an ARA headset and hear information about interesting places. If the user allows, also commercials and special offers from the shops he/she is passing by could be sent to the headset. This kind of guide application could be also implemented to museums, exhibitions and shopping malls.

2.5.4 Calendar

The Calendar application has been implemented in previous research for example in [10, 20]. The idea in the 3-D audio calendar application is that the user could record calendar notes to the system and place them to some direction in the space. For example 9 A.M. would be on the left side of the user, noon in front and 3 P.M on the right side. In this application, no head-tracking is required since the calendar is supposed to move along with the motion of the head.

2.5.5 Driving Assistant in a Car

One of the obvious places where there already is a loudspeaker setup is a car. This setup could be utilized in head-tracking. In this application the user would wear an ARA headset which would be connected directly or through an ARA mixer to a PDA or a mobile phone. The user would hear the natural sound environment of the car and could communicate with passengers and hear the important binaural cues of the engine and the traffic.

In addition to this, the PDA or mobile phone could play back music or the radio to the ARA headset. The loudspeakers of the vehicle would be used as static anchor sources for head-tracking. The anchor signals could be played back from a special CD or the mobile phone or the PDA could have a radio transmitter, which sends the anchor signals at some radio frequency. The radio of the car would be tuned to this radio frequency and the anchor signals could be easily played back. This would not require any additional hardware to the vehicle and all the necessary signal processing would be done by the PDA or mobile phone. It is a sort of extended hands-free system which most mobile phones already have.

This application could be used to improve safe driving. For example, if the driver looks too long at other direction than the road, the head-tracker would notice this and the system would send a warning to the ARA headset and could, for example, wake up the driver. Because most PDAs and mobile phones have GPS, also navigation instructions and other location based information (commercials of gas stations and malls, tourist information, traffic information, weather forecasts,...) could be offered to the user though the ARA headset. Of course incoming calls could be answered with the same system.

Chapter 3

Headset and Mixer Development

3.1 Sound Reproduction by Headphones

3.1.1 Headphone Acoustics

Sound reproduction with headphones is very different from reproduction with traditional loudspeakers. The main reason for this is that the signal route after the loudspeaker is completely removed. This includes the effect of the listening room and binaural cues caused by the listener's body, head, torso, and ears [23]. The signal route is replaced by a much smaller but yet not a simple system, the ear. This system becomes even less predictable, because the ear canal responses have large individual differences [30]. This is a problem in designing headphones, since finding a frequency response, which is applicable for most people, is demanding.

Desired Frequency Response of Headphones

The design criteria for loudspeaker and headphone frequency responses are different when the purpose is to reproduce sound signals as unaltered and natural as possible. Usually, a high-fidelity loudspeaker's desired frequency response is flat when measured in free-field (in an anechoic chamber). Loudspeaker reproduction is the primary target when recording and mixing commercial production material. This is problematic for the design process of headphones, since headphones replace the whole setup. Usually, the motivation for the design process of a headphone is to try to make the listening experience somewhat similar to the experience with loudspeakers. In such cases, the same frequency response should be heard at the eardrum as in the case of a loudspeaker and a flat frequency response is not the target [23].

The free-field sound pressure level (SPL) of a loudspeaker reproduces the SPL at the microphone in the live sound field. However, the body, torso, head and ears of the listener affect the frequency response, and the ear signal response is not flat anymore. Loudspeakers can be compared based on their free-field frequency responses since the listener would have also disturbed the sound field in the same manner in the live sound field. However, in the case of headphones it is difficult to produce the same kind of ear signal response. This kind of approach has led to the concept of free-field calibrated headphones. The free-field calibrated headphone's frequency response tries to replicate the ear signals for a loudspeaker in front of the listener. However, the frontal localization is not completely achieved, which colors the response. The response of the headphone is often a compromise between this distorted response and a completely flat one and mostly a matter of taste [4]. Because of this inaccuracy, exact recommendations for the frequency response are difficult to form. However, some guidelines are presented for example in [23].

In contrast to the free-field calibrated headphone, another approach is the diffuse-field calibration. This approach emphasizes that in normal listening environments, direct sound contributes only to a fraction of the total sound entering the ear. If the distance of the listener to the loudspeaker is larger than the reverberation radius, the total sound field consists mainly of reflections. In [23], the design criteria are presented for both free-field and diffuse-field calibration. Figure 3.1 shows an average of the two different approaches and presents the desired frequency response for a headphone in a "combined field" [23]. Usually, the frequency response has a peak around 3-4 kHz because this way the headphone should sound natural. The reason for the use of this peak is the ear canal effect, which does not occur when a headphone is placed on the ear, see Equation (3.1) [30].

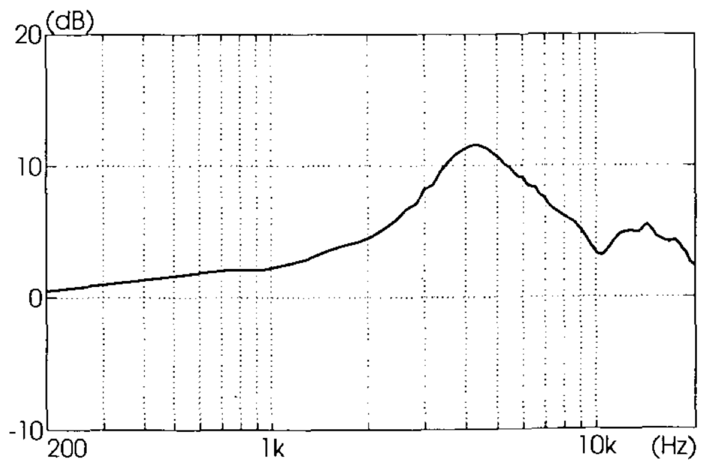

Figure 3.1: The design goal for a frequency response of a "combined field" headphone (after [23]).

Headphone frequency response design goals presented in earlier researches have mainly concerned more traditional types of headphones (supra- and circum-aural). However, the types of headphones that are of more interest for ARA applications are in-ear headphones (for different headphone types, see 3.1.2). These headphone types have some differences in the desired frequency responses. With supra- and circum-aural headphones, the ear canal is mostly open and in the case of in-ear headphones the ear canal is closed (which is not the natural situation for the ear). This difference affects the ear canal resonances. The ear canal can be considered a rigid tube which ends to a relatively hard ear drum with a frequency-dependent acoustic impedance. The normal length of an ear canal is about 22.5 mm and the diameter is about 7.5 mm. In the case of an open ear canal, the canal acts as a quarter wavelength resonator with one end being closed by the ear drum and the other end open to the air [13]. A cross-sectional view of the ear can be seen in Figure 3.2.

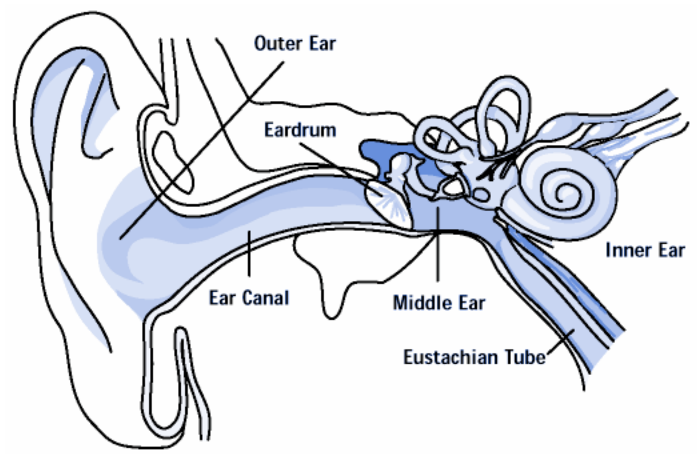

Figure 3.2: The cross-sectional view of the human ear (after [24]).

For an open ear canal, the first ear canal resonance can be calculated with

$$f_0 = \frac{c}{4l}, \tag{3.1}$$

where c is the speed of sound and l is the length of the ear canal (in reality, the effective length is longer because of attached mass effect and should be taken into account). Using Equation (3.1), the lowest resonance for a 22.5 mm long ear canal occurs at about 3.8 kHz (normally it is around 2-4 kHz because of the attached mass effect). This is the case when there is nothing blocking the ear canal. When a headphone blocks the ear canal, this resonance disappears, and the sound field would appear unnatural unless it was taken into account in designing the frequency response. Normally, this effect has been taken into account, as shown in Figure 3.1.

The frequency response changes a lot when an in-ear headphone closes the entrance of the ear canal. In this case, the ear canal is closed from both ends and it starts to act more like a half-wavelength resonator. If we again assume the ear canal to be a rigid tube with hard walls in both ends, the lowest resonance frequency can be calculated with

$$f_0 = \frac{c}{2l}, \tag{3.2}$$

where c is the speed of sound and l is the length of the ear canal. For an 22.5 mm long ear canal, the lowest half-wave resonance occurs at about 7.6 kHz, but because of variations in the length of the ear canal and the design of the earplug, this resonance can be somewhere between 5 and 10 kHz. The quarter-wave resonance is often taken into account in the frequency response by designing a peak around 2-4 kHz, however, the new and unnatural half wave resonance is almost always present. This resonance should be compensated in order to make the perceptual response of the headphone sound flat and uncolored. The problem with this resonance is, however, that it is more challenging to predict than the quarter-wave resonance, because the fitting of the earplug varies and ear canals are individually very different. The half-wave resonance is also much steeper, and because of this the compensation dip should also be quite sharp. Thus, the position at the frequency scale becomes even more important.

The effect of the half wave resonance (the peak just after 10 kHz) can be seen in Figure 3.3. The curve is the difference between the response measured at the ear canal simulator and the response at the matched-impedance tube (for measurement techniques see, Section 3.2). It is also possible that the peak after 10 kHz is a multiple of the peak at about 5.5 kHz, which might be the first resonance. It is possible that the transducer has it's own resonance almost at the same frequency as the half-wave resonance (as can be seen from Fig. 3.4) and the half-wave resonance is therefore not seen very clearly before the second multiple after 10 kHz.

The earplug can be inserted in many ways depending on the size and shape of the ear canal. When the same user puts the headphones on many times, some variation also exists. These occasional variations along with the fact that the ear canal is not ideal (like assumed), probably lead to the simplified design of the headphone's frequency response, where only the quarter-wave resonance is compensated. However, the headphones can be equalized individually after measuring them (see Section 3.5 for further details). In reality, the ear canal is not a straight tube and the ear drum impedance is frequency-dependent, thus attenuating and spreading the resonance formed in the ear canal, making it more difficult to predict and model [22, 37].

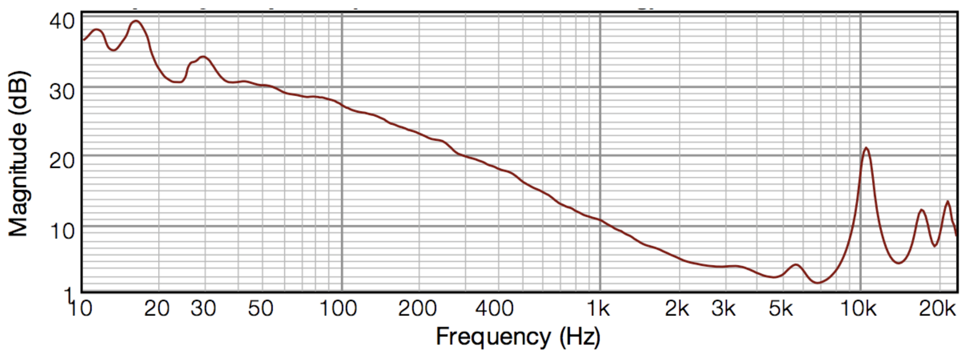

Figure 3.3: The difference between frequency response of the Philips SHN2500 transducer in the ear-simulator and in the matched-impedance tube.

The Pressure Chamber Effect

In headphone listening, the reproduction of low frequencies is somewhat less demanding than in the case of loudspeakers. This is mainly due to the so-called pressure chamber effect [4]. When listening to loudspeakers, the sound field is relatively large (typical room volume is tens of square meters) and the sound pressure produced is spread to this space. However, when listening with headphones, the volume of the sound field is reduced to a very small quantity. In the case of traditional circum-aural headphones, the volume is about 30 cm^2, but when using insert-type of in-ear headphones, it can be as small as 1 cm^2 [4]. It is clear that in the small cavity much higher sound pressure levels can be produced, because the ear can be considered a pressure detector. When the wavelength of sound is much larger than the dimensions of the headphone, the amplitude and phase of the sound pressure can be considered evenly distributed in the cavity, but there should not be massive leaks present (see Chapter 3.1.1). In reality, there are always some leaks, but they can be minimized with good fitting.

When the pressure is in phase with the volume displacement of the earphone's transducer and its amplitude is proportional to it, the adiabatic compression law gives

$$pV^{1.4} = C, \tag{3.3}$$

where p is the pressure (sound pressure and atmospheric pressure), V is the volume of the cavity, and C is a constant. The sound pressure per unit volume displacement is

$$\frac{dp}{dV} = -\frac{1.4 * C}{V^{2.4}}, \tag{3.4}$$

and it is constant for a given volume. Equation (3.4) is frequency-independent, which is not the case with loudspeaker sound reproduction. Because sound pressure produced by a loudspeaker is proportional to the volume acceleration of the membrane (not to the volume displacement like in headphones), the loudspeaker requires larger a displacement to produce low frequencies [4].

Figure 3.4 presents the pressure chamber effect and includes three frequency responses of the Philips SHN2500, which are measured in the ear canal simulator, in the matched-impedance tube and in free-field. It can be noticed that the bass response attenuates respectively. Because the pressure chamber effect is present in the ear canal simulator, the amplification of low frequencies is strongest in these measurements. The same effect can also be found in Figure 3.3 (the transfer function of the ear canal simulator).

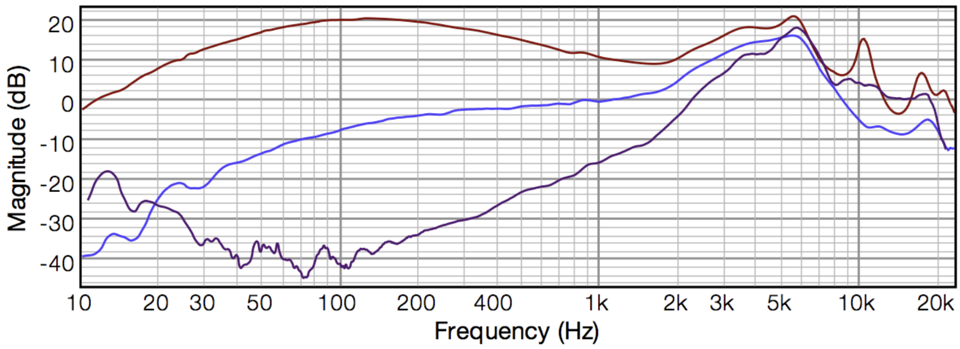

Figure 3.4: Frequency responses of the Philips SHN2500 transducer: the highest curve is measured in the ear canal simulator, the middle curve is measured in the matched-impedance tube, and the lowest curve is measured in free-field.

Leakage and Conduction of Sound

In headphone listening, the leakage around and through the headphone is one of the most important factors influencing the low and mid-range frequency reproduction. The reason is that the pressure chamber effect does not occur unless the system is properly sealed [4]. Some leaks are always present, however, and to control the leaks, some of the leaks can be deliberate parts of the acoustical design (meaning that the leaks are intentional) [26].

Sound can leak in two directions: from the headset to the environment and from the environment to the headset (here the headset refers to the whole system, including headphones and ears). The first case is important from the pressure chamber perspective. The second one is important in ARA applications, since the sound leaking from the real environment sums in the ear canal with the pseudoacoustic representation produced by the transducer. This summing causes coloration and deteriorates the pseudoacoustic experience [9].

Sound can leak through many different paths. In traditional circum- and supra-aural headphones, the sound can leak through the headphone itself, through the porous cushion and through leaks between the cushion and the skin. Unwanted leaks are also often caused by hair between the skin and the cushion [26]. Because the leak is variable, its influence to the frequency response is also variable.

The leakage is somewhat different in the case of in-ear headphones. Especially in the case of insert-type of headphones, the leakage can be very minimal if the earplug is fitted properly. On-ear type of earplugs, which are loosely placed on the ear canal entrance, were used in the older ARA headsets. The fitting was not very tight and changed much, depending on the size and shape of the ear canal. It could also be inserted on the ear in many different ways affecting the repeatability of the response [32]. Insert-type of headphones (like Philips SHN2500, see Chapter 3.4.1) are fitted more tightly and consistently. Figure 3.5 shows the different leakage routes of an in-ear headphone.

Figure 3.5: Leakage paths of an in-ear headphone (after [33]).

No matter which type of in-ear headphones we are dealing with (on-ear or insert), the leakage routes are the same, even though the degree of the leaking effect varies. Severe leakages can be caused by openings between the earpiece and the skin. These openings can be considered acoustically as a one wide and thin slit surrounding the casing or as several narrower slits in parallel [33]. Such leakage is often present if the headphone is made of a stiff or hard material like plastic. One way to overcome this problem is to use porous or foam-like material (sometimes cloth), to cover the casing. In some insert-type of headphones, rubber-like material is also used, giving good sealing and fitting [33].

Another route for the sound to leak is through bone and tissue, but this type of conduction is attenuated very strongly [3], and the leakage around and through the headphone is much more prominent with loosely-fit in-ear headphones. In the case of insert-type of headphones, the conduction is more severe. Most in-ear headphones also contain some openings in the back of the casing, from which sound can leak outwards through the headphone. The holes can be for example covered with clothing inside the enclosure or bass tubes. The human skin and flesh are also elastic, which leads to vibration of the earpiece. The vibration of the headphone causes motion also in the ear canal, which is interpreted as sound by the ear [33]. This is more of a problem with tightly fitted insert headphones where the coupling is efficient.

Equivalent circuits

Headphones can be modeled with simple equivalent circuits. In this study only moving-coil type of transducers are considered, since most in-ear headphones are based on this technology. Other technologies used with in-ear headphones include piezoelectric and tube transducers. A simplified schematic view of a moving-coil transducer can be found in Figure 3.6. The voice coil, which is in a static magnetic field, is in contact with the membrain, and when the current in the voice coil changes, the magnetic field moves the coil and the membrane accordingly [4].

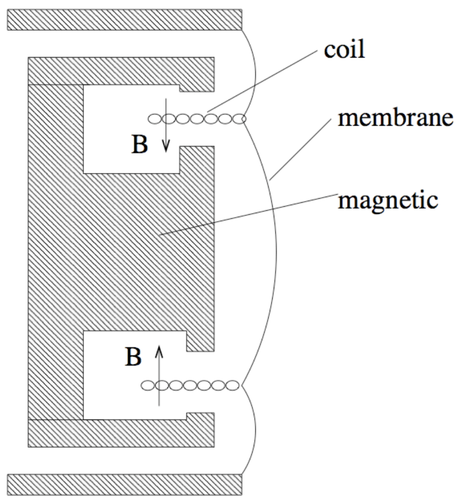

Figure 3.6: A schematic view of a moving-coil transducer (modified from [4]).

In headphone modeling, there are three different domains which have to be considered. These domains are electric, mechanical and acoustical. Complex mechanical and acoustical systems are modeled in equivalent circuits using simpler electrical circuits, consisting of inductors, resistors, and capacitors. Headphones have the basic acoustical components, which are volume compliance (C which is connected to the ground), porous element (R, L in series) and compliant membrane (R, L, C in series). The pressure p is equivalent to the voltage U, and the volume velocity q is equivalent to the current I [4]. The interconnection between the different domains consists of converting the electric sound signals (the electric domain) into membrane movement (the mechanical domain), which creates sound pressure waves (the acoustical domain). To connect these different domains in one circuit diagram, there is a need to use special transformers and gyrators. Figure 3.7 shows the equivalent circuit of a moving-coil headphone, where the three different domains are kept separated, but connected with a gyrator (upper part where electric domain connects to mechanical domain) and with two transformers S_1 and S_2 connecting mechanical to acoustic domain. Z_{load} is the acoustic load after the headphone (ear canal etc.).

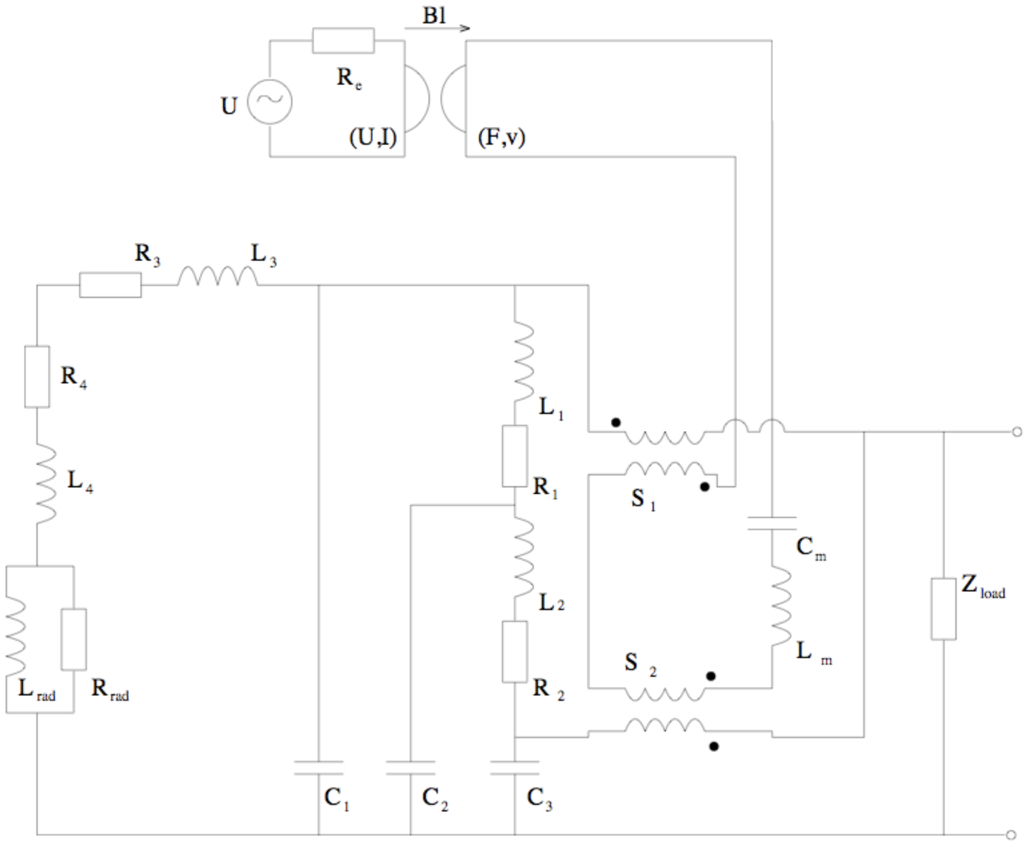

Figure 3.7: Equivalent circuit of a moving-coil headphone (modified from [4]).

3.1.2 Headphone Types

Many types of headphones are commercially available. The most traditional ones are placed either resting on the earlobe (supra-aural headphones) or around the whole ear like a cup (circum-aural headphones). Much research has been done concerning these two main types (for example [23, 4]). However, modern portable devices often come (and are used) with small, in-ear headphones because of their lightness, size, good usability and outward appearance. There is little reliable research material available in this area, yet. In-ear headphones are also the most obvious choice for ARA applications because of the usability reasons. Another reason supporting this choice is that the in-ear earplugs interfere less with the binaural signals captured by the integrated microphones than in the case of larger supra- or circum-aural headphones [9]. The insert-type of headphones have also developed a lot in recent years. Earlier models (also used in previous ARA headsets) were loosely-fit in-ear headphones [32]. They are placed loosely on the entrance of the ear canal and thus do not seal the ear canal completely. This results in leakage around the headset. The leaked sound and the electrically transmitted sound are summed in the ear canal, which causes coloration, and the pseudoacoustic representation is deteriorated when used as an ARA headset [33].

Fortunately, new types of in-ear headphones have emerged in the market, which are better for making the ARA headset acoustically transparent. These earphones are more tightly fitted to the ear canal entrance sealing it effectively. This way the leakage is reduced considerably. In fact, they resemble earplugs for hearing protection and even hearing aids very closely [33]. When considering the insert-type of headphone's acoustic properties, many acoustic characteristics of earplugs (see [12, 11]) and hearing aids (see [7, 16]) are applicable to some extent and can be used as background material for further studies. Different headphone types are shown in Figure 3.8.

Figure 3.8: Different headphone types from left to right: circum-aural, supra-aural and in-ear headphones.

One new and interesting group of headphones are the bone conducting headphones. Rather than playing the sound to the listener's ears, they transform the sound signals into mechanical vibration. They utilize the conduction of sound to the inner ear through the bones of the skull. The headset is located in front of the ear and it rests on the head having straight connection to the cranial bones. There are some advantages in this kind of headphones. For example, they do not cover the ear, so it is possible to listen to the natural sound environment at the same time while wearing the headphones. This would be very beneficial in ARA applications. There might occur some cross-talk between the ears, since the sound could conduct from one side to the other through bone and flesh, which is not desired.

3.2 Measurement Techniques for Headphones

There are many ways to measure the frequency response of headphones. The most common one is the artificial ear, which tries to replicate the human ear. Sometimes the measurements are also done with a dummy head (like the Brüel&Kjær Head And Torso Simulator Type 4128 with a middle ear simulator Type 4158). In both of the techniques, microphones are placed either at the end of the ear canal (where the ear drum would be) or at the entrance of the ear canal. There are many opinions, which method is better, and also the type of headphone, which is under test sets some restrictions. When the microphone is at the end of the canal, also ear canal resonances are recorded. A different curve is obtained, when the frequency response is measured at the ear canal entrance, especially when the canal is blocked because then the effect of the ear canal is absent. It is also possible to make the measurements in a freefield without any kind of artificial ears or heads, but then the bass response is attenuated, because the pressure chamber effect does not occur (see Chapter 3.1.1).

In this study, a specially constructed ear canal simulator, a matched-impedance tube, and free-field measurements are used, which are described in their own subsections.

3.2.1 The Ear Canal Simulator

Some measurements presented in this thesis are made with an ear canal simulator. The in-ear headphones can be inserted into the simulator in the same manner as in a real ear. The ear canal simulator is a 25 mm long plastic tube with a diameter of 9 mm (about the same size as an average human ear canal). The headphone is inserted to the other end of the tube and the other end is blocked by a hard wall. A Sennheiser KE-4 electret microphone is tightly fitted through a hole in the wall. Figure 3.9 shows the ear canal simulator and a headphone inserted to it.

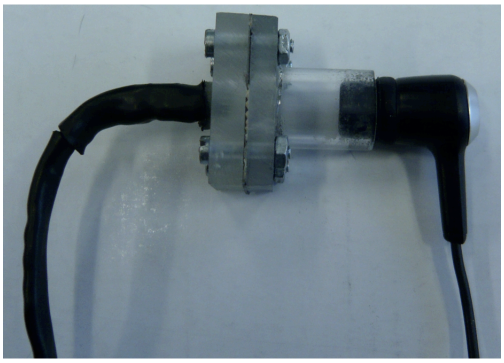

Figure 3.9: The ear canal simulator and a headphone.

This setup replicates the real situation quite well, since the ear canal is blocked by the headphone and the microphone is at the ear drum position. Its downside is that the simulator is made of hard plastic, while in reality the ear canal is softer and the ear drum's impedance is frequency-dependent. The effect of the tube is, however, close to the effect of the ear canal (see Chapter 3.1.1). This can be seen at low frequencies as amplification (pressure chamber effect) and also at higher frequencies as half-wave resonances (at about 10 kHz in this case, see Equation (3.2)). When the responses recorded with the ear canal simulator are compared to the responses of the other measurements, the effects are clearly seen.

3.2.2 The Matched-Impedance Tube

To get rid of the reflections and resonances of the ear canal simulator, also another tube was constructed. This plastic tube is 12 m long and has a diameter of 9 mm (about the same diameter as the human ear canal). The headphone is inserted to the other end of the tube and the other end is plugged with soft wool to eliminate the reflection from the end. A Sennheiser KE-4 electret microphone is tightly fitted through a hole in the wall of the tube, 3 cm from the end where the headphone is. From these measurements, same kind of response can be obtained as from the ones made with the ear canal simulator, except for ear canal resonances and amplification at low frequencies. Figure 3.10 shows the matched impedance tube with a microphone and a headphone inserted to it.

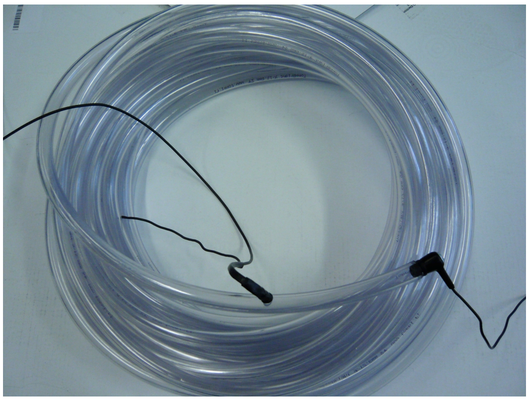

Figure 3.10: The matched impedance tube with a microphone and a headphone inserted to it.

3.2.3 Free-field Measurements

A measurement was also conducted in the anechoic chamber. The headphone was fitted through a hole in a large baffle (just like in loudspeaker element measurements) and the impulse response was recorded 19 cm from the baffle with a Røde NT2-A studio microphone (it was chosen because it has a good sensitivity and low noise at low frequencies). Mid and high frequencies were reproduced quite similarly as in the case of the matched-impedance tube (and ear canal simulator, but without the ear canal resonances). Low frequencies were almost completely attenuated, because there is no cavity where the pressure chamber effect could occur.

3.3 Previous ARA Headsets

There have been a number of ARA headsets in the previous project. Most of them have been constructed from loosely-fit in-ear headphones by integrating microphones to them. Their acoustical properties and sound quality have been studied in articles [32] and [33], for example. The basic design goal for an ARA headset is that it should produce a good enough copy of the surrounding sound environment, so that the user could wear the headset for longer periods of time.

One of the problems with the previous headsets was that their output properties relied heavily on the fitting of the headphone. This makes the reproduction of low frequencies difficult, since leaks have low-pass like characteristics. Also the pressure chamber effect might not work and the leakages are uncontrolled and hard to predict. Another problem is the occlusion effect, which means that when the ear canal is blocked, the own voice of the user sounds boomy and hollow, [32]. This effect is, however, present also in the new type of headsets and even more pronounced because they seal the ear canal more tightly.

It is clear that to obtain the best possible sound quality of the headphone, some equalization of the magnitude response is needed. It is possible to make average equalization curves based on measurements from many test persons. If individual equalization curves are used, the sound quality of a headset should be improved even more efficiently [32].

As the leaked sound sums with the sound produced by the earphone in the ear canal, the result is a colored version of the real sound environment. If the leakage paths are known, this can be compensated with signal processing [33] or equalization (see Chapter 3.5).

3.4 Headset Development

Since there are no commercial ARA headsets available, the headsets have to be constructed as described earlier. One interesting group of products has emerged: small, in-ear, active noise canceling headphones. The advantage over normal in-ear headphones is that they already have integrated microphones in them. Earlier, the microphones had to be somehow installed to the headphones, which led to problems concerning for example the usability of the ARA headset. There has also been a change in the in-ear type of headphones from loosely-fit headsets towards tightly-fit, insert-type of headphones. These insert-type of headphones resemble certain types of earplugs for hearing protection and also hearing aids [33].

The basic idea behind the noise canceling headphones is that they normally have a small control box, which has signal processing electronics to reduce unwanted external noise. In principle, it should enhance the listening experience and the volume of the headset could be lower than in the case of no noise canceling, since the signal-to-noise ratio is higher. The signal from the microphones is routed to the control box. The control box recognizes noise signals and sends them to the headphone in an inverse phase. This way the unwanted noise signal will be reduced in the ear canal. Even when the noise canceling is turned off, the earplug attenuates outside noises quite well because of tight fitting.

In the case of an ARA headset, we do not use the original control box nor the signal processing of the noise canceling headset anyhow. The earpieces, including the microphones and the earphone transducers, are disconnected from the control box and connected to an ARA mixer designed in the project.

3.4.1 Philips SHN2500

One of the interesting new noise canceling headphones is the Philips SHN2500, which is a small in-ear headphone. It is an earplug, which seals the ear canal quite tightly. Before constructing any ARA headsets using them, the properties of the integrated microphones and the transducers of the Philips SHN2500 were measured and tested. The headset can be seen in Figure 3.11.

Figure 3.11: Philips SHN2500 Headset and the Control Unit (after [31]).

The Microphone of Philips SHN2500

For the measurements, the noise canceling control unit was disconnected from the headset and the microphone was connected to a biasing network. The network blocked the DC voltage from the captured signal. This filtering was done with a resistor of 2 kΩ and a capacitor of 1 μF with a 1.5 V voltage source. The measurement was done in an anechoic chamber. A sine sweep was excited from a Genelec 8030A active monitor and recorded with the microphone of the Philips headset in three different angles: 0, 45, and 90 degrees.

The frequency range of the Genelec loudspeaker is 66 Hz - 20 kHz ($\pm$2.5 dB) which is enough for our purposes, since the most interesting part of the microphone‘s magnitude response is at high frequencies (the acoustic head-tracking system uses a frequency range above 17kHz). A high-quality Brüel&Kjær 4191 free-field microphone was used as a reference.

The frequency responses of Genelec 8030A source loudspeaker recorded by the Brüel&Kjær and the Philips microphone are shown in Figure 3.12. The responses are very similar up to about 15 kHz. After that, the frequency response of the Philips microphone drops somewhat faster than the Brüel&Kjær's, but not too dramatically. After all, there is not much useful audio information at frequencies above this. However, if anchor sound sources are used for acoustic positioning and head-tracking, this property of the microphone has to be kept in mind, since often these sound sources are placed at quite high frequency bands (like 17-20kHz). But it should not be a problem, only a matter of gain adjustment (see Chapter 2.4).

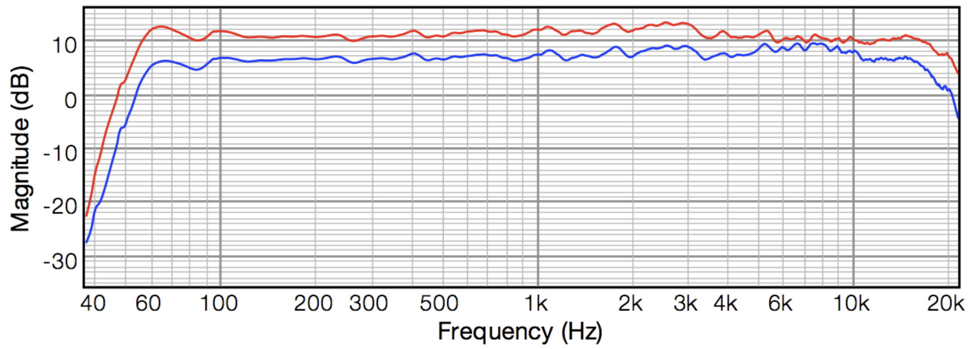

Figure 3.12: Frequency response of Genelec 8030A recorded with Philips SHN2500 microphone (lower curve) and Brüel&Kjær 4191 free-field microphone (upper curve).

The directivity of the microphone starts to increase above about 3.5 kHz but has a controlled behaviour as can be seen from Figure 3.13. The reason for increased directivity at higher frequencies, when measured off-axis, is that the more the wavelength of the excitation signal approaches the dimensions of the earphone itself, the more the earphone shadows the signal. The higher the frequency, the more directive the microphone is.

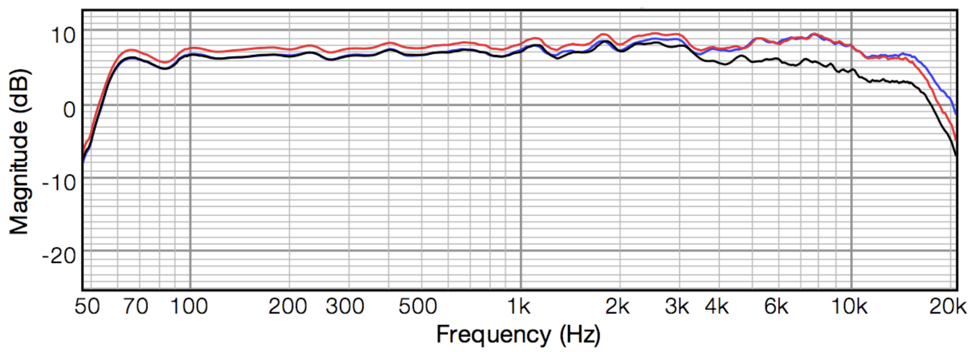

Figure 3.13: Frequency response of Genelec 8030A measured with Philips SHN2500 microphone in three different angles: blue is 0 degrees, red is 45 degrees, and black is 90 degrees.

The Earphone Transducer of Philips SHN2500

Another feature that has to be considered when choosing the ARA headphone is the frequency response of the earphone. In the case of normal loudspeakers, it is often desired to have a frequency response as flat as possible when measured in a free-field. This is not the case with earphones since when they are placed on the listener's head, a very complex signal route is formed as described earlier in Section 3.1 and [23].

The Philips SHN2500 headset's transducers were measured in three different setups as described in Section 3.2. The most interesting case is the ear canal simulator, which is the highest (red) curve in Figure 3.4. The most visible thing in the frequency response to be noticed is the strong amplification of lower frequencies when the earplug is placed in the ear canal simulator. This can be explained by the impedance matching of the earphone to the tube and also with the pressure chamber effect as described earlier in Section 3.1.

There are also strong resonances present in the response. They occur above 5 kHz and the first one (at about 5.5 kHz) is a design feature of the transducer (it can be seen in all of the three responses measured in different setups). The resonance is slightly more pronounced in the ear canal simulator, and it is possible that the first half-wave resonance of the closed cavity happens to be at same frequency range. The next peaks and dips in the ear canal simulator response are caused by standing waves in the closed cavity (the ear canal is a half wave resonator now, see Section 3.1.1).

From the free-field measurement (the lowest curve in Figure 3.4) it is also possible to find the same resonance at about 5.5 kHz as from the other measurements, which is characteristic of Philips SHN2500. What is also worth noticing, is the lack of low frequencies when measured in free-field. The small transducer cannot produce enough sound pressure to reproduce low frequencies without the help of a transmission channel like the ear canal, [4]. The frequency response curve's shape is, although, close to the design goal presented in Figure 3.1. The analysis of the sound quality of the new headsets based on empirical user test are described in Section 5.1.

3.5 ARA Mixer Development

The previous ARA mixers have been simple mixers with no equalization properties. They have only had microphone preamplifiers and headphone amplifiers. They had also one input connector for signals to be included in the pseudoacoustic environment and one output for transferring binaural recordings captured by the microphones. The new mixers developed in this work have the same basic properties as the previous ones, but with analog equalization. There is a simple analog, first-order high-pass filter to compensate the frequency response at low frequencies. This is needed because often the insert-type headphones have too strong bass reproduction, which also varies depending on the fitting of the earpiece to the ear canal. The fitting is important, since it affects the leakages especially at low frequencies. The filter is constructed with a simple RC-circuit consisting of a resistor, a potentiometer, and a capacitor. The circuit for one channel is shown in Figure 3.14. A potentiometer (R2 in Figure 3.14) is used in series with a resistor (R1). This way the cut-off frequency can be changed according to the amount of leakage. The cut-off frequency of the high-pass filter is

$$f = \frac{1}{2\pi RC}, \tag{3.5}$$

where R is the resistance of the potentiometer and the resistor (R=$R1$+$R2$) and C is the capacitance of the capacitor. The values for the components were chosen so that the cut-off frequency can be varied from 6 to 720 Hz (-3 dB).

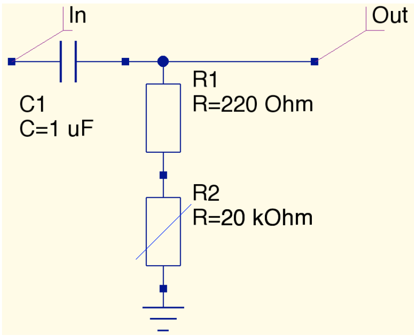

Figure 3.14: The high-pass filter circuit (R2 is a potentiometer with maximum resistance of 20 kΩ).

The effect of the filter can be seen in Figure 3.16, which shows frequency responses of Philips SHN2500 measured in the ear canal simulator with different cut-off frequencies. The lower the curve, the higher the cut-off frequency. The lowest curve is measured with the highest cut-off frequency; 720 Hz (-3 dB) and this way the response is quite flat. The high-pass filter and the other equalization controls do not alter the binaural microphone signals, which are routed to the output connector of the ARA mixer (and possibly sent to a distant user). The equalization is only done to the binaural signals routed back to the user's own ears and this way makes the pseudoacoustic environment more realistic.

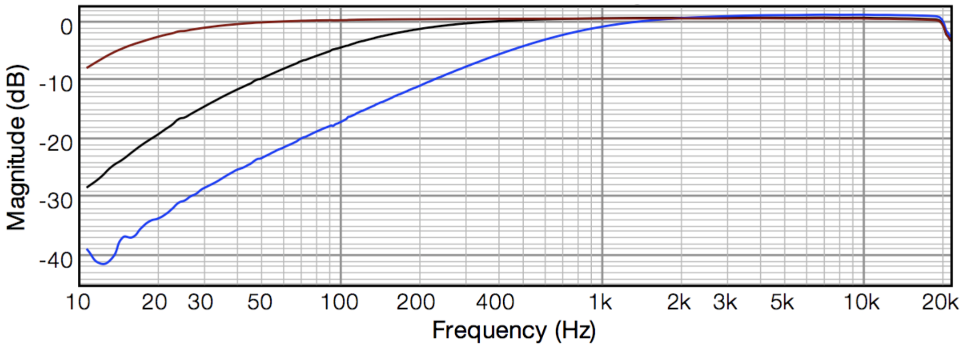

Figure 3.15: Frequency response of the ARA mixer with three different high-pass filter settings (bass control).

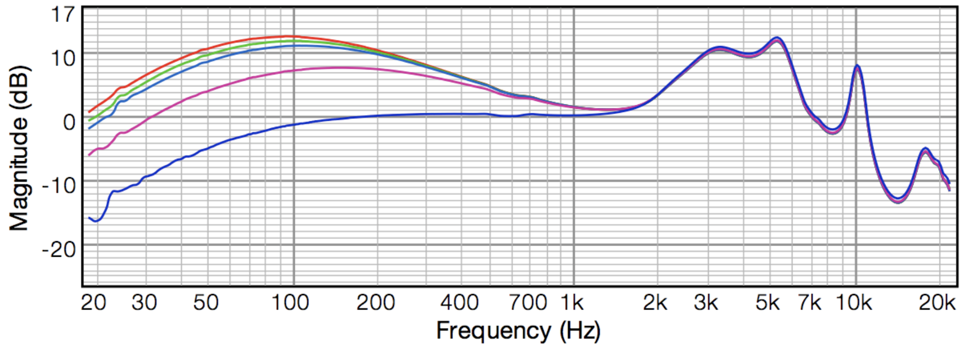

Figure 3.16: Frequency response of Philips SHN2500 measured in the ear canal simulator with three different high-pass filter settings (the effect of the bass control).

There are also two adjustable parametric equalization sections to compensate the resonance problems described in Section 3.1.1 (the circuits are based on [15]). The center frequencies of the resonators can be moved and the Q-values and gain are also adjustable. The parametric equalization is based on an active LRC-filter, where L (inductor) is constructed with a gyrator. Figure 3.17 shows the block diagram of the ARA mixer.

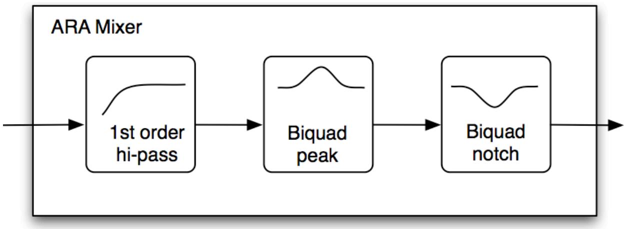

Figure 3.17: The block diagram of the ARA mixer.

The Q-value determines the bandwidth and the steepness of the peak or notch. The higher the Q-value, the narrower and steeper the peak or notch is. The first parametric equalization control is supposed to be used as a peak. It is there to make sure that the effect of the natural quarter-wave resonance occurs even though the ear canal is occluded. It can be moved in the range of 700-3200 Hz.

The other equalization control is supposed to be used as a notch, which compensates for the half-wave resonance that occurs when the ear canal is closed (a standing wave in the closed cavity). The notch can be moved in the range of 1.8-8.5 kHz. The cut-off frequency of the high-pass filter for the bass control can be moved from 6 Hz to 720 Hz (-3 dB). It is also possible to turn the equalization completely off.

The equalization circuit for the boost and the notch for one channel is shown in Figure 3.18. There are two branches (inside the dashed lines), which are connected to the operational amplifier U1b through a 10 kΩ potentiometer and the other end is connected to the ground. The left branch controls the peak in the range of 700 - 3200 kHz and the right branch controls the notch in the range of 1.8 - 8.5 kHz. The 10 kΩ potentiometer's setting determines wether it produces a peak or a notch, and in the center there is no effect. When the potentiometer is dialed to the input side, it shunts the input to the ground and produces a notch in the frequency response at this resonant frequency. When it is dialed to the inverting input of the operational amplifier, it shunts the feedback to the ground and produces a peak. There is no effect when the potentiometer is directly in the center, because the LC-filter's effect is equal on both inverting (-) and non-inverting (+) inputs.

To be able to change the Q-value and make the peak or the notch narrower or wider, a 10 kΩ potentiometer ("Resonance 10 K" in Figure 3.18) is used between the capacitor (C1 or C3) and the simulated inductor (inside the dashed lines). If the LC-filter has a non-zero resistance, it no longer looks like a short to ground at the center frequency, it looks rather like a resistor (at this frequency). This effect makes the frequency selectivity less radical and lowers the Q-value. On the contrary, when the potentiometer is set to its minimum, the Q-value is the highest.

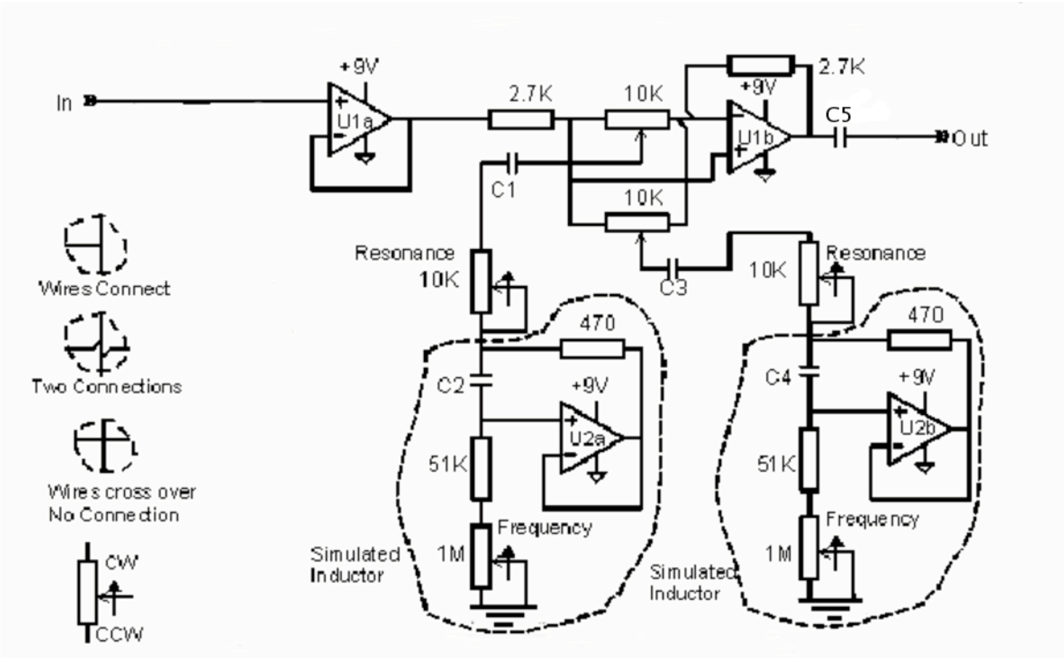

Figure 3.18: The equalization circuit of the ARA mixer (modified from [15]).

It is characteristic of a series LC-filter that it has a high impedance except at its resonant frequency, where the impedance drops to a minimum. The resonant frequency can be calculated with

$$f = \frac{1}{2\pi\sqrt{LC_1}}, \tag{3.6}$$

where L is the simulated inductance and C_1 is the capacitor value. The simulated inductance L can be calculated with

$$L = 470\Omega * R * C_2, \tag{3.7}$$

where R is the series resistance (that is R=51 kΩ + R_x, where R_x is the resistance of the 10 MΩ potentiometer) and C_2 is the capacitor value. The values for the capacitors chosen for the ARA mixer are shown in Table 3.1.

To be able to change the resonant frequency and move the peak or notch in the frequency domain, the value of the capacitor (C) or the value of the inductor (L) has to be made variable. Because variable capacitors and inductors are bulky and expensive, simulated inductors are used. The simulated inductors are shown in Figure 3.18 inside the dashed lines and they consist of an operational amplifier, two resistors, a capacitor and a potentiometer. Generally this kind of circuit is called a gyrator. The 1 MΩ potentiometer's position determines the value of L and thus also the resonant frequency.

With this kind of ARA mixer, it is possible to make the headset even more acoustically

Table 3.1: Capacitor values for the parametric equalization (Figure 3.18).

Capacitor	Value
C_1:	33 nF
C_2:	3.3 nF
C_3:	15 nF
C_4:	1 nF
C_5:	22 μF

transparent. The equalization properties can be also turned completely off. Figure 3.19 shows the effect of a single parametric equalization at the range of 700 - 3200 Hz and Figure 3.20 shows the effect at the range of 1.8 - 8.5 kHz. In these frequency responses the gain and the Q value are set to their full values. This way it is possible to get amplification of about 8-16 dB depending on the center frequency.

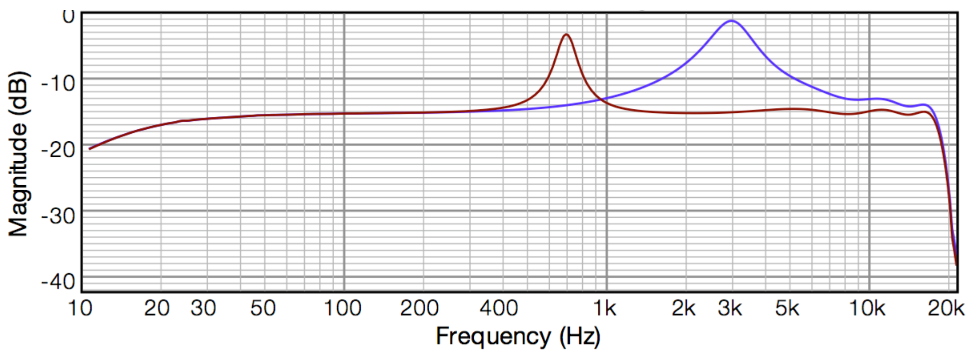

Figure 3.19: Frequency responses of the ARA mixer with a single equalization peak set to it's border frequencies 700 Hz and 3200 Hz.

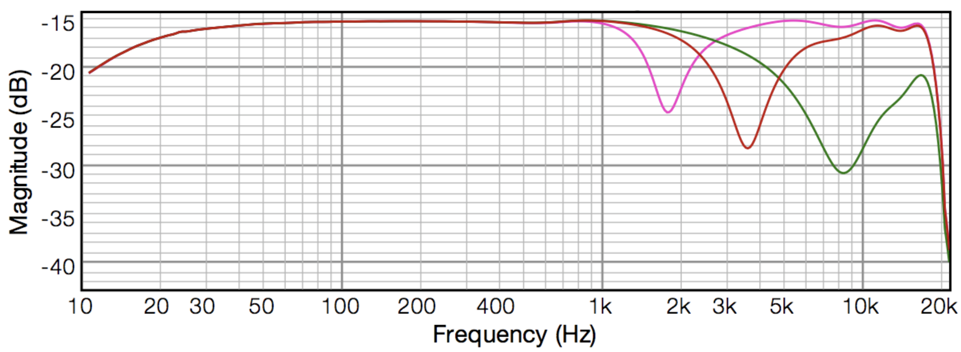

Figure 3.20: Frequency responses of the ARA mixer with a single equalization notch set to it's border frequencies (1.8 kHz and 8.5 kHz) and in the center (3.5 kHz).

3.5.1 Connectors and Controls

There are six miniplug (3.5 mm) connectors and four controls in the mixer, which are also shown in Figures C.1 and 3.21.

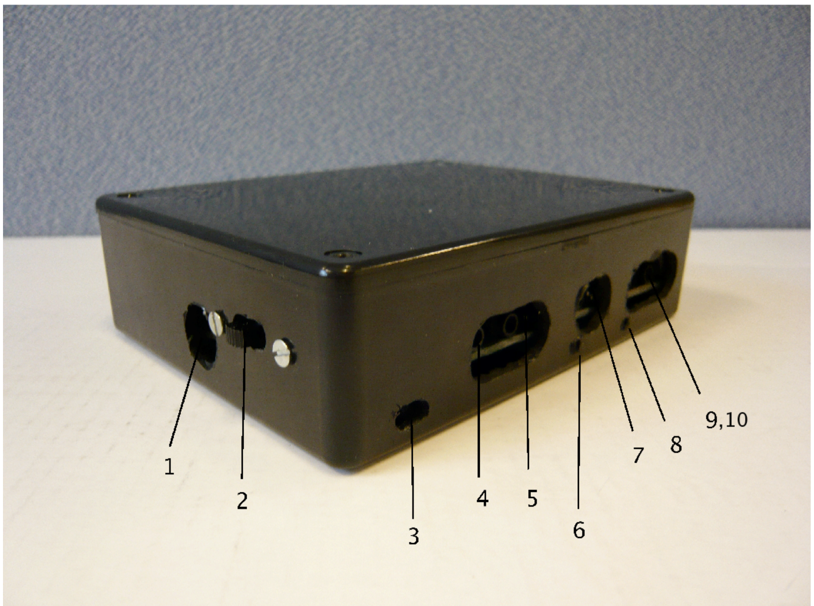

Figure 3.21: The ARA mixer

1. "Eq out" is the output connector for an external equalization device (the only connector at the end of the box). From this connector, unequalized binaural microphone signals are obtained when the connection is activated by the slide switch.

2. Power switch: in the left position the power is off

3. Equalization switch: leftmost position is for external equalization ("Eq out" connector is activated), in the middle there is flat frequency response (Eq off) and in the rightmost position equalization with bass control is on (Eq on).

4. "Out" is the connector from which the unequalized microphone signals can be routed to a computer or some other device.

5. "Mic" is the connector for the microphones

6. Level for right channel

7. "Phones" is the headphone connector

8. Level for left channel

9. "In 1" is the connector for an input signal from a computer or some other device. The equalization does not affect these signals.

10. "In 2" is the input connector for an external equalization device

Actually "In1" and "In 2" are identical and also "Out" and "Eq out" are identical, except "Eq out" is activated with a slide switch.

At the end of the box (next to the "Eq out" connector), there is a power switch. When the switch is turned to the left side (in the direction of the "Eq out" connector) the power is turned off. There are two 9 V batteries, which should last about 20 hours of use.

At the side where most of the connectors are, there is a slide switch (number 3, next to "Out " connector). The switch has three positions and allows the user to select the signal route from the microphone to the earphone (so it effects the pseudoacoustic representation). The three possibilities from left to right are:

- External equalization ("Eq out" connector is activated)
- Flat frequency response (Eq off)
- Parametric equalization with bass control (Eq on)

When the external equalization is turned on (the leftmost position), the microphone signal can be routed from the "Eq out" connector to an external equalization device. After the equalization, the signal should be routed back to the mixer's "In 2" (or "In 1" since they are identical). In normal circumstances, the switch should be turned to the rightmost position (Eq on).

The volume controls are constructed with multiturn potentiometers (one for each channel), so they are very accurate. More amplification is obtained by turning the screw clockwise with a small screwdriver through the hole in the box. The volume control for the right channel is between the "Mic" and "Phones" connectors and for the left channel the control is between the "Phones" and "In 1" connectors.

3.5.2 Circuit and Component Values of ARA Mixer

The complete circuit for one channel is shown in appendix Figure C.1 and the other channel is identical to this one. The main volume control of the ARA mixer is in the headphone amplifier, which is actually an attenuator. Resistors, which have a line crossing over them, are trimmers with adjustable resistance. Either TL072 or TL074 type of operational amplifiers can be used.

The component values of the circuit are shown in appendix Tables C.1, C.2, and C.3. The voltage source for the circuit consists of two 9V batteries. One of them provides +9V positive voltage and the other one -9V negative voltage to the circuit. Ground (0V) is obtained by connecting the negative pole of the first one and the positive pole of the latter one.

This kind of ARA mixer is flexible and makes the ARA headset acoustically more transparent. One of the advantages of the mixer is its very low latency in signal processing because it is implemented as an analog circuit. Commonly used digital signal processing hardware and software cause delays of at least a few milliseconds, which is unacceptable for augmented reality audio applications, and deteriorate the pseudoacoustic representation. Even fractions of a millisecond of latency in signal processing can be disturbing. This deterioration is caused by comb filter effect when the pseudoacoustic representation is delayed compared to the leaked sound. The two sound fields sum in the ear canal and because of the comb filter effect, some frequencies are attenuated and some pronounced.

The best results are obtained when thorough measurements are conducted, and the mixer is equalized individually for each user and headset. The reason for this is that there are differences in the ear canal resonances and fitting of the earphone between users. However, it is not always possible to conduct proper measurements due to lack of time or facilities and hardware. In this case, it is possible to find equalization settings, which are generically suitable for most people. This can be done by collecting a large enough database of representable measurements from a set of users and equalize the mixer based on these findings [27, 36]. The results of the equalization based on measurements and empirical user test are described in Section 5.1.

Chapter 4

Acoustic Positioning and Platform Realization

4.1 Platform

As described in Chapter 1, there is a need for a platform where ARA applications and ARA technology could be developed and tested. We built a setup in a room for this purpose but the development is still ongoing as this thesis is written. The objective was to build a platform with support for the connectivity of the basic ARA components (headsets), point-to-point communication with binaural signals, and possibly support by video communications and display. The platform should also support user positioning for position-aware application experiments. It should also be easy to test demonstrations developed outside the project.

The platform was constructed to a room with dimensions of 6 X 9 m. The hardware for the platform includes a computer, internet connection, soundcard, amplifier for head-tracking anchors, and six anchor sources installed to the walls of the room. Also wireless headphone receiver-transmitters are available for the ARA headsets in order to ensure easier mobility.

A Macintosh computer with an 8-channel soundcard was connected to a 6-channel amplifier, which is used to control the anchor sources. There is a 2-way communication software installed in the computer, so that binaural telephony and other possible applications can be tested between two places.

The software can send and receive full-quality stereo audio over UDP protocol. It can also send and receive some side information over TCP protocol. This can be used for example for exchanging tracking data between two places. Figure 4.1 shows a screenshot from the 2-way communication software, which is implemented with Pure Data software (PD). PD is a real-time graphical programming environment for audio, video, and graphical processing [29].

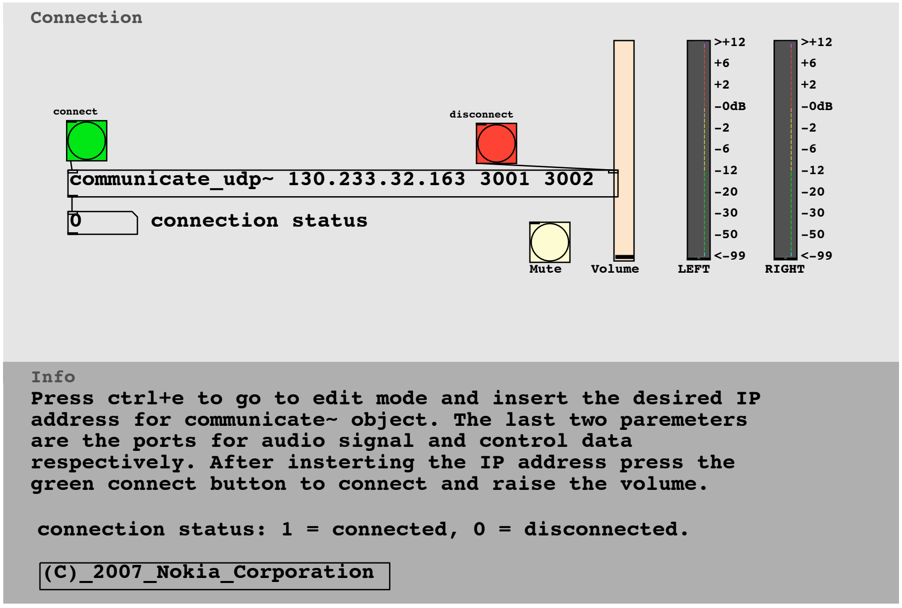

Figure 4.1: A screenshot of the 2-way communication software (courtesy of Nokia Research Center).

The original idea was that the platform would be ready by the end of this phase of the project. The plan was to have a working acoustic head-tracking system and a two-way communication software with the possibility to send positioning coordinates and orientation information to an another location (TML). On the TML side, they already have a video-based head-tracking system. However, the goal turned out to be somewhat laborious and by the end of writing this thesis we have the communication software working and the anchor sources installed to the walls but the head-tracking algorithm is still in progress. Figure 4.2 shows the original diagram of the platform.

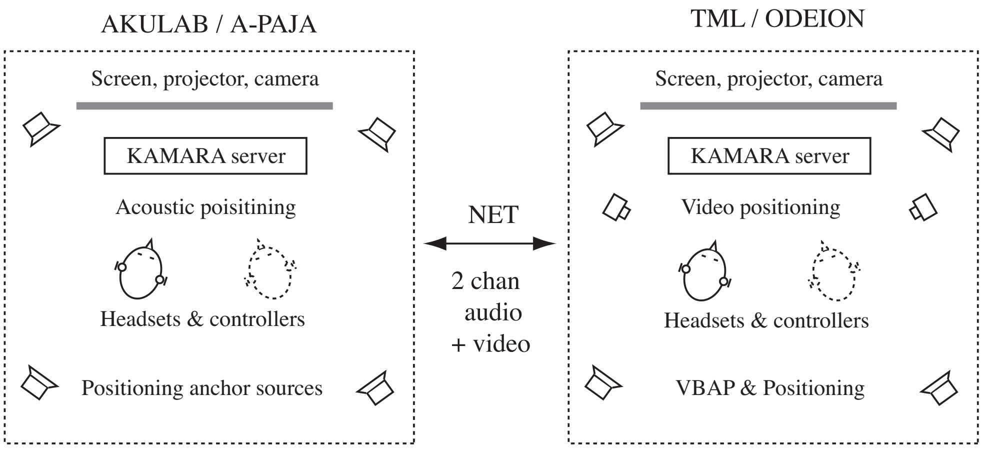

Figure 4.2: System diagram of the platform.

The anchor sources (tweeters) were installed to the walls of the room. Figure 4.3 shows the floor plan of the room and the six tweeter positions on the walls (the unit of measurement is centimeters).

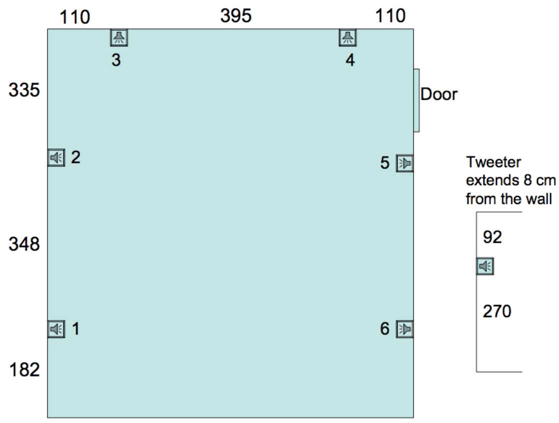

Figure 4.3: Floor plan of the platform room showing the anchor sources on the walls.

4.2 Acoustic Positioning and Head-tracking

In order to make authentic augmented reality audio applications, it is necessary to keep track of the user's movement and orientation. This enables the system to create static or moving sound objects to the user's sound environment in a controlled manner, even when the user is moving in the space. The requirements for the system are low power consumption, light weight, no wires, and comfort.

With these requirements and available technology, it was decided that acoustic positioning and head-tracking based on binaural signals and known anchor sources would be used in the implemented platform. There has also been prior research and implementation in the Laboratory of Acoustics and Signal Processing on this kind of tracking [34, 14, 35].

4.2.1 The Binaural Method

The basic idea behind the binaural positioning method used in this study is that static anchor sources emit some known reference signal. These signals are captured by the microphones in the user's ear canal entrance. By processing the signals, the orientation of the subject and distance to the anchor sources can be calculated. To make the system less complex, anchor locations, signals, and signaling times are known. This kind of synchronous tracking allows the exact distance and orientation in relation to each anchor to be solved, but some synchronization is needed [35]. In an asynchronous case, the signaling times are not known.

The positioning is done by investigating the differences in the two microphone signals. As described earlier in Chapters 2.3.2 and 2.3.1, the most important differences are the interaural level difference (ILD) and the interaural time difference (ITD). The ITDs are calculated with cross-correlation between the known source signal and the microphone signal. If the anchor signals are not known, the cross-correlation is computed between the two microphone signals. In a synchronous case, three anchor sources have to be known in order to place the target anywhere in the space. If only two anchors are used, position in a half plane divided by the anchors can be obtained [35].

4.2.2 The Anchors and Anchor Signals

High Frequency Anchor Signals

To obtain better results by using several anchor sources, there is a need for separating the sound sources from one another after recording them with binaural microphones. There are many ways for doing this [34, 14, 35]. One of them is to divide the frequency scale into as many sub-bands as there are anchor sources. Each anchor plays it's own sub-band and after they are recorded, they can be separated.

The reference signal can be almost anything from music to speech or noise. The idea, however, is that the user should not hear the reference signal separately, on the contrary it should be masked by the background noise. This could be difficult to implement adaptively, so other kind of solution has to be applied. One way is to utilize the fact that human auditory sensitivity decreases at low and high frequencies and by placing the reference signals at these frequencies could be beneficial [14].

Other factors supporting the use of a high-frequency reference signal is the room acoustics and the nature of background noise. Reflections caused by the room and the furniture and also the normal type of office background noise have a low-pass type of behavior. Also the user's own voice decays towards high frequencies, so there is less background noise at high frequencies and good signal-to-noise ratio is easier to obtain. However, the shadowing effect of the user's head is stronger at high frequencies, which could be a problem if the reference signal is intended to be completely hidden to the background noise. Then the signal-to-noise ratio gets weaker and ITD is harder to obtain at high frequencies because of strong shadowing. Despite all, it is justified to use high frequencies for reference signals [14].

When using only high frequencies as reference signals, there are not so much demands for the anchor sources (loudspeakers). This is because often low frequencies are more difficult to reproduce and they require a larger cabinet for the loudspeaker. If only high frequencies are considered, very small loudspeakers can be used. For example, one possible option is to use headphone transducers. The problem we encountered with this kind of anchors is that they do not produce enough SPL to be captured by the ARA headset's miniature microphones in a large room. It would have been beneficial to use these small headphone transducers as anchors since they are not too directive at high frequencies because of their relatively small diameter compared to the wavelength of the signals. In smaller areas, these kinds of transducers could be used.

The anchor sources were chosen to be loudspeaker tweeters, since they are enough omnidirectional and can produce the required SPL levels without any problems. It was decided to use six static anchor sources together with a six-channel amplifier. The amplifier is a Rotel RB-976, which can produce 60 W for each channel. The tweeters were selected based on their low directionality properties. The choice was SEAS TAFD/G, which is a 19 mm High Fidelity dome tweeter with a smooth, extended frequency response. The tweeter is shown in Figure 4.4 and its frequency responses are shown in Figure 4.5. The responses are measured in an anechoic chamber with a Brüel&Kjær 4191 free-field microphone in four different angles.

Modulated High Frequencies as Anchor Signals

Although there are many benefits supporting the use of high frequencies as anchor signals, there are also drawbacks. The shadowing effect of the user's head increases towards high frequencies and also other objects start to shadow as well. This can interfere the tracking process. One way to overcome this would be to use ultrasounds above 20 kHz with higher levels since they are not heard by people, but the hardware (microphones and loudspeakers) limit this possibility. Their sensitivity normally decreases strongly above 20 kHz or even earlier.

Figure 4.4: SEAS TAFD/G tweeter used for positioning and head-tracking.

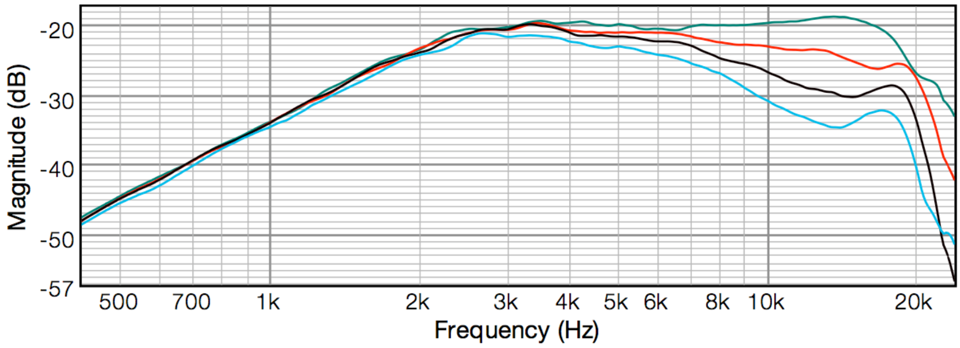

Figure 4.5: The frequency responses of SEAS TAFD/G in four different angles: 0, 30, 60, and 90 degrees.

One possible realization for anchor signals has been presented in [14] and later in [35]. This solution utilizes also high frequencies, but in a more sophisticated way. In this method a low-passed (about 1-2 kHz) reference signal modulates different carrier frequencies. Every anchor has it's own carrier frequency and they play the same reference signal modulated to a different frequency. This happens synchronously. The original reference signal can be obtained from the recorded signal by demodulating the signal with the appropriate carrier frequency [35].

The benefits of this approach are coherence between the anchor signals and also computational savings. Because the same reference signal is played synchronously through the different anchors, the time differences between the arriving signals from different anchors can be calculated even if the reference signal is not known. This happens by computing the cross-correlation between two demodulated anchor signals. The computational saving comes from the lower sampling frequency, which can be used because the reference signal is a low-passed signal.

4.2.3 The Positioning and Head-tracking Processes

The positioning and head-tracking can be divided into smaller processes. Different kinds of information have to be computed from the microphone signals. The user's orientation and distance in relation to the anchors have to be obtained in order to find out the exact location where the user is and also the direction he/she is facing. In the following sections, the basics of the computation processes are explained.

Orientation

Orientation means the direction the user is facing. In our setup, there are six anchor sources emitting a known reference signal. The signal is divided into six sub-bands with the help on filters $P_1(f) - P_6(f)$. The orientation can be obtained by detecting the ITD values. The following shows how this is done to one anchor. The anchor signal from the *i*th anchor is

$$X_{ref,i}(f) = X_{ref}(f)P_i(f), \tag{4.1}$$

where $X_{ref}(f)$ represents the fourier transform of the reference signal and $P_i(f)$ is the filter, which divides the signal in sub-bands.

After this, the cross-correlation has to be calculated between the reference and recorded signals.

$$R_{l,ref,i} = \int_{\tau-T}^{\tau} \Phi_i(f)S_{l,ref,i}(f)e^{j2\pi f\tau}df, \tag{4.2}$$

where

$$S_{l,ref,i}(f) = E\{X_1(f)X_{ref,i}(f)^*\} \tag{4.3}$$

is a sampled cross-spectrum between left channel and the reference. The weighting function is

$$\Phi_i(f) = \frac{P_i(f)}{|S_1(f)S_{ref,i}(f)|^\gamma}, \tag{4.4}$$

where $P_i(f)$ represents a frequency mask for the *i*th source and is also used for filtering the anchor signal. γ is a parameter, which affects the amount of magnitude normalization.

The estimation for the ITD is the distance of the maximum values of $R_{r,ref,i}$ (right channel) and $R_{l,ref,i}$ (left channel) [34].

$$ITD_i = \max(R_{l,ref,i}) - \max(R_{r,ref,i}). \tag{4.5}$$

Now that the ITD is known, the angle ϕ_i of the head relative to the *i*th anchor can be solved by assuming an an ITD model, such as delay

$$d_{i,ITD} = R_{head}\{\phi_i + \sin(\phi_i)\}, \tag{4.6}$$

where R_{head} is the radius of the user's head. The lateral angle for the other anchors can be calculated in the same manner.

Distance

If we use synchronized anchor signals, the distance is easy to estimate by turning the traveling time from the source to the microphone into distance. If there is no information on when the signal was emitted (asynchronous case), the initial value of the time delay estimate *D* sets a reference point to which the upcoming delays can be compared to [35].

As the user moves, the maximum values in the cross-correlation responses also move. The change in distance relative to the *i*th source can be obtained from the average movement of the maximum values by

$$d_i = 0.5(\max(R_{l,ref,i}) - \max(R_{r,ref,i})) - D, \tag{4.7}$$

where *D* is the initial value of d_i.

After the distances and angles to all the anchor sources are calculated, the exact position and orientation of the user can be obtained from the information with geometrical calculations. The dimensions of the room and the coordinates of the anchor sources have to be known. Also the height of the user's ears from the ground should be known. As the ITDs and angles are known, the orientation and position of the user can be calculated. The algorithm is still in the development phase at the point of writing this thesis.

4.2.4 Using Modulated Anchor Signals

If coherently modulated anchor signals are used, each anchor can be solved by demodulation. The modulated signal played by the *i*th anchor is

$$x_{i,ref}(t) = x_{ref}(t)\cos(2\pi f_i t), \tag{4.8}$$

where $x_{ref}(t)$ is the low-passed reference signal, which modulates a carrier signal $\cos(2\pi f_i t)$. To resolve the reference signal, the recorded signal has to be demodulated with the corresponding carrier frequency [35].

Synchronization

As the reference signal played by the ith anchor is n samples long, the tracking system buffers the first n samples when started. Then a cross-correlation is calculated between the ith reference signal and the other channel of the recorded signal $s_r(t)$ or $s_l(t)$. After this, the reference signal in the memory is circularly shifted to synchronize the system with the recorded signal. After synchronization, the recorded frame should match a frame from the reference signal.

The synchronization degrades as the user moves. After a while, the recorded frame and the frame from the reference signal do not correlate enough anymore, and the system has to be synchronized again [35].

4.2.5 The Tracking System for the Platform

As described earlier, the head-tracking system for the platform is work under progress. The methods presented in the previous sections only provide one possible solution. It still not fixed what kind of anchor signals and positioning algorithms will finally be used in the platform. Preliminary measurements, tests and earlier research have, however, shown that the system has the potential for this kind of tracking and research continues after writing this thesis.

Chapter 5

System Testing and Analysis

5.1 Mixer and Headset Listening Tests and Analysis

The ARA mixer and headset (Philips SHN2500, see 3.4.1) were evaluated through listening tests [27] and a usability study [36]. The listening tests were conducted with individual and non-individual equalization as well as with no equalization at all. The equalization was based on measurements described in the next chapters. The usability of the ARA headset was evaluated in real-life situations with a set of subjects. A group of testees were given an ARA headset and mixer and they were instructed to wear the headset for longer periods of time in everyday-life situations. The main question for the testees was: *"Would you be willing to wear this kind of a headset in everyday-life situations?"*.

5.1.1 Measurements for Determining Equalization Parameters

To find proper equalization settings, ear canal transfer function measurements were conducted for four test subjects. The measurements were performed in the anechoic chamber at the Department of Signal Processing and Acoustics in Helsinki University of Technology. The goal of the measurements was to determine the resonances and the transfer functions of the ear canal in different cases. The measurements were conducted with a microphone inside the ear canal. The first measurement was done with an open ear canal (the natural hearing situation) and in the second measurement the sound was routed through the ARA headset (the pseudoacoustic representation). In the latter case, the headset blocked the ear canal entrance. These cases were compared to each other and the equalization was done based on the comparison. The goal was to make the pseudoacoustic representation as similar to the natural situation as possible. A non-individual equalization setting based on all the measurements was also calculated [27].

The Measurement Setup

The setup for the ear canal transfer function measurements was the following. A Genelec 1030A loudspeaker was placed 2.5 m in front of the test subject. The loudspeaker was at the height of the subject's head. FuzzMeasure software [28] with logarithmic sweep technique [8] was applied. The sweep was played with the loudspeaker and recorded with a microphone inside the ear canal. This way it was possible to obtain the transfer function of the transmission path, including the loudspeaker and the measurement microphone.

A Knowles Acoustics FG-3329-P07 miniature in-ear microphone was located inside the ear canal of the subject (about 5 mm from the ear canal entrance). The transfer function was first recorded from an open ear canal. This represents the natural listening experience. After this measurement, the microphone was attached to an ARA headset. The wires of the microphone were passed between the rubber cushion and the frame of the headset. The cushion was further sealed to prevent leaking. Figure 5.1 shows the miniature microphone attached to the headset.

Figure 5.1: Knowles Acoustics FG-3329-P07 miniature in-ear microphone attached to the Philips headset.

When the subject wore the headset, the miniature microphone was inside the closed ear canal approximately in the same position as in the case of an open ear canal. The ARA headset was connected to the ARA mixer (the equalization was turned off at this point). This measurement represents the unequalized pseudoacoustic representation of the excitation signal inside the closed ear canal. In both measurements, the idea was to find out the transfer function from outside the ear to the microphone inside the ear canal and how it changes when an ARA headset is used.

Figure 5.2 shows the results of the first measurement phase. It represents two transfer functions recorded from one person: the transfer function from a loudspeaker to an open ear canal and also to the ear canal when the sound goes through an unequalized ARA headset (pseudoacoustic representation). The black line is the open ear canal case.

The most important feature is the lowest ear canal resonance at about 2 kHz. The gray line is the case, where the user is wearing an ARA headset with no equalization and the microphone is in the closed ear canal recording the pseudoacoustic representation of the sine sweep.

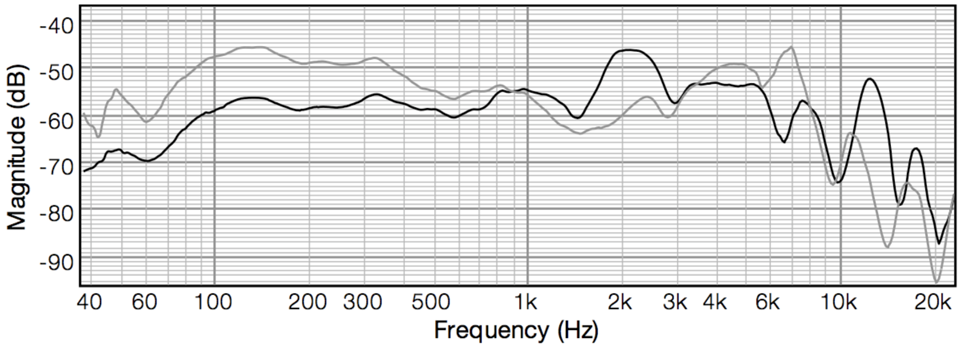

Figure 5.2: Example of two transfer functions from source loudspeaker to a microphone inside the ear canal. Black line represents an open ear canal case and gray line is the case where the sound is played through an unequalized ARA headset.

The natural quarter-wave resonance is attenuated in the pseudoacoustic case and there is a new peak at around 7 kHz caused by the half-wave resonance. Also low frequencies are pronounced in the pseudoacoustic representation. The attenuation of high frequencies (above 10 kHz) can be explained by the directivity properties of the microphone of the ARA headset (see Fig. 3.13). The loudspeaker was in front of the test subject so the binaural microphone was in an angle of more than 90 degrees from the loudspeaker. The dimensions of the headset are close to the wavelength of the excitation signal at high frequencies so the headset starts to shadow the incoming wavefront causing attenuation. This effect is not so pronounced in normal listening environments, where there are reflections from all surfaces (the measurements were done in an anechoic chamber).

Measurement results

After the first measurement round, the open ear canal transfer function and the pseudoacoustic transfer function were known. Individual target equalization curves could be obtained by taking the difference of the two functions on the dB scale and after this, the curve was adjusted to the mixer by hand and through measurements. Also a generic equalization curve was generated from the average of the four individual equalization curves (Figure 5.4). Two individual equalization curves are shown in Figure 5.3. They are measured from the mixer and adjusted by hand based on the measurements described earlier. The curves have peaks and a notches at slightly different frequencies because of varying ear canal lengths.

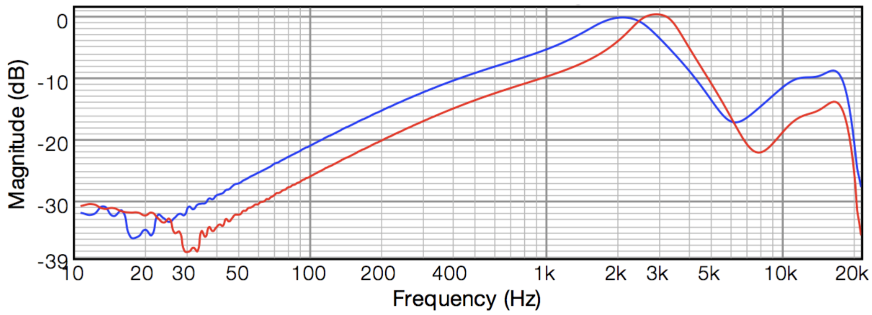

Figure 5.3: Two individual equalization curves measured from the ARA mixer.

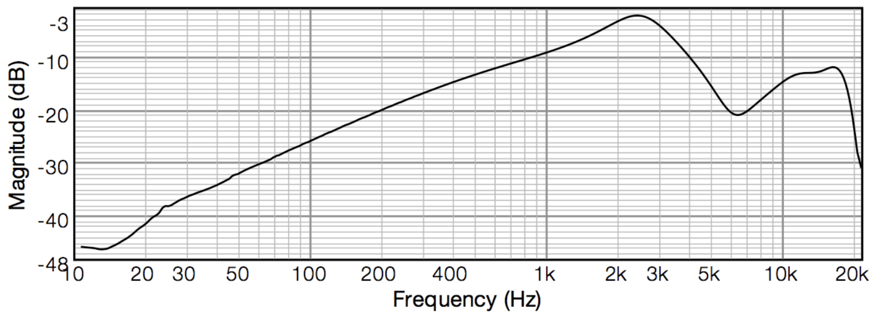

Figure 5.4: Generic equalization curve based on four individual curves (measured from the mixer).

Figure 5.5 shows the results of the second measurement phase, where the effect of the individual equalization was tested. The measurement data shows a clear improvement in the pseudoacoustic representation compared to the unequalized case shown in Figure 5.2. Low frequencies are very close to the natural situation and the resonances are almost compensated.

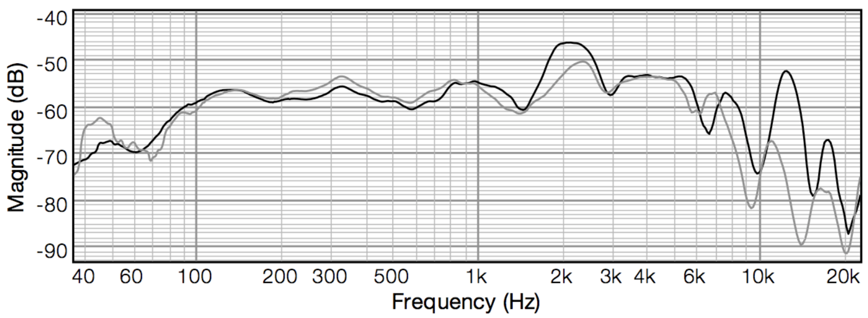

Figure 5.5: Transfer functions from source loudspeaker to a microphone inside the ear canal. This corresponds to Fig. 5.2 except now the gray line is the case where the sound is played through an individually equalized ARA headset.

5.1.2 Listening Tests

A small preliminary listening test was arranged with four test subjects. The goal was to find out whether the overall sound quality of an ARA headset could be improved with the simple analog equalization of the ARA mixer. Another goal was to study how a generic (non-indivual) equalization filter would perform compared to individual equalization. The equalization was based on the ear canal transfer function measurements described earlier. All the test subjects were experienced listeners [27].

Listening Test Setup

The listening tests were conducted in a standard listening room with a stereo loudspeaker setup. At first the user did not wear an ARA headset and listened to the sample with unoccluded ears. Then he wore an ARA headset with no equalization and listened to the sample again. After this, he had to evaluate the unequalized pseudoacoustic representation compared to the natural hearing experience. The same evaluation was also repeated with the individual and the average equalization settings (both compared to the natural hearing situation). The subjects did not know which equalization setting they were listening to (no equalization, individual equalization or a non-individual equalization setting).

The subject was allowed to put the headsets on and off as many times as necessary for the evaluation. The subjects were instructed to ignore any noise or level differences. It was also instructed that the users inserted the earplugs properly to the ear canal entrance in order to avoid leakage. Furthermore, the users were told to make the evaluation based on the assumption that they would be wearing the headset in everyday life for longer periods of time.

The evaluation was done with three different samples; acoustic music, male speech, and pink noise. The length of the music sample was 24 seconds and it was played in a continuous loop. The speech sample was a looped 5 second sentence of a male voice (in Finnish) and the pink noise was continuous. The A-weighted sound pressure level of the sound samples was approximately 70 dB. Between the samples, the order of the equalization settings was changed randomly. There were two listening rounds to get more reliable results and the listening test took about 30 minutes.

The evaluation was done using the standard DMOS-scale (Degredation Mean Opinion Score), where the test subject assesses the aberration of a sample compared to a reference sample. In this test the pseudoacoustic representations were compared to the natural hearing experience, which was the reference. The DMOS scale is from 1 to 5, where 5 is the best grade (imperceptible difference) and 1 is the worst (very annoying). Table 5.1 shows the DMOS scale [5]. The test subjects were also instructed to give comments and opinions about the overall sound quality of the different cases.

Table 5.1: DMOS grading scale (after [5]).

Grade	Impairment
5	Imperceptible
4	Perceptible, but not annoying
3	Slightly annoying
2	Annoying
1	Very annoying

Listening Test Results

Both individually and non-individually equalized settings made a clear improvement compared to a non-equalized setting [27]. The overall results are shown in the boxplot of Figure 5.6, where the left bar is for the non-equalized, middle bar for the individually equalized and the right bar for the non-individual equalization setting. The boxes in Figure 5.6 have lines at the lower quartile, median, and upper quartile values. The whiskers are lines extending from each end of the boxes to show the extent of the rest of the data. The results are normalized according to ITU-R BS.1284-1 recommendation [5].

As can be seen from Figure 5.6, the non-equalized pseudoacoustic representation was given an average grade of 3, which means that it was slightly annoying compared to the natural listening situation. Both of the equalized cases got an average grade of 4, which means that compared to the natural listening experience, the difference was perceptible, but not annoying. Because the number of subjects was only four, no conclusions can be drawn on which one of the equalizations, individual or generic, is better. The equalization was done based on only one measurement round (as described in Section 5.1.1), so no fine-tuning was made to the equalization curves.

The test subjects' comments about the non-equalized pseudoacoustic representation were that low frequencies were highly pronounced and the representation did not sound very natural. For two of the four test subjects, the individually equalized setting worked very well and for the other two, the generic equalization curve seemed to be slightly better. However, the differences were considered very small and highly dependent on the fitting of the earplug to the ear canal entrance [27].

5.1.3 ARA Mixer and Headset Usability Evaluation

We also arranged a preliminary usability test with four subjects. The idea was to collect user experiences about the headset and mixer from real-life situations. The main question

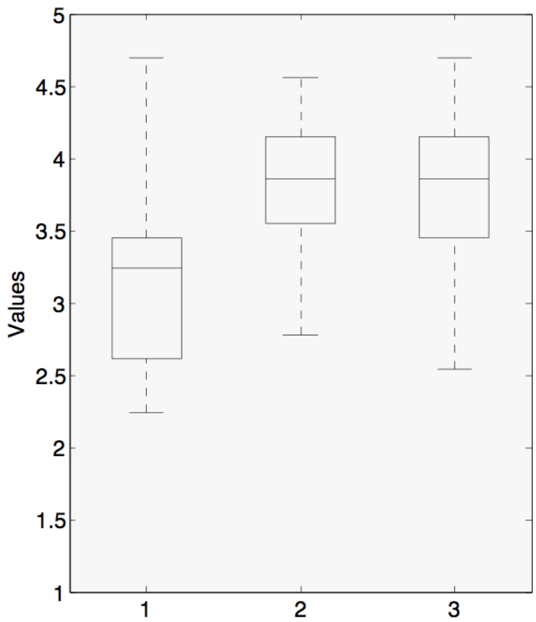

Figure 5.6: Results of the listening test: 1 = non-equalized, 2 = individually equalized, 3 = non-individually equalized.

for the subjects was that would they be willing to wear the headset for long periods of time in normal everyday life if useful applications were available. The subjects were instructed to make notes in a diary on everything they considered good or bad about the headset and mixer combination [36]. They were provided with a non-individually or individually equlized ARA headset (Philips SHN2500) and a mixer (shown in Fig. 5.7) for a couple of days and were encouraged to use the headset as much as possible in different situations.

The testees made notes in a diary for example about the following things:

- Coloration caused by the headset
- Feel of space
- Feel of direction
- Irritation caused by the headset to the ear canal
- Comfort or discomfort caused by the headset
- The amount of noise and disturbances in the sound
- Limitations and effects caused by the wires

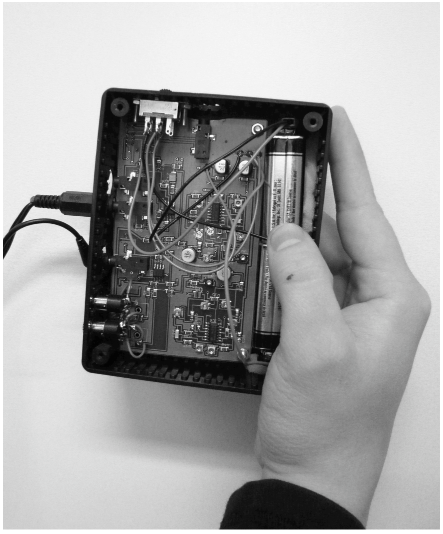

Figure 5.7: Prototype ARA mixer.

The test subjects wore the headset in many different situations. These situations included for example:

- Dining
- Speaking
- Walking
- Running
- Driving a car or sitting in a bus or a tram
- Listening to music in home and in a car
- Watching television
- Speaking to a mobile phone

The subjects could use their imagination in selecting the use cases and comment on anything they thought was relevant. Also new application scenarios were welcome. The users wore the headset for about 20-30 hours in sessions from a few minutes to eight hours.

Although many people are used to wear earplugs or headphones daily when listening to music, there are still few, if any studies, on how people would perceive and accept an ARA headset, especially for longer periods of time. One of the most important factors in acceptability of this kind of headset is the sound quality, which should not differ too much from natural acoustics [27]. Another key issue is portability. However, since people are accustomed to carry mobile phones, mp3-players and PDA's, it should be no problem to carry a small ARA mixer also. In addition to general usability of an ARA headset, the goal was also to gather information on different aspects that have effect on the usability of such headsets.

The overall comments from the testees were very positive. Especially the sound quality was perceived very good and applicable for many everyday-life situations. The most critique raised from the wires, which produce mechanical noises to the headset and they tend to be in the way all the time. The following chapters summarize and comment the diary reports from the testees. A more compact summary of the diaries is shown in Table B.1 [36].

Sound Quality

The results and comments about the sound quality of the equalized headset were very positive [36]. The testees said that they quickly got used to the headset and after a while they did not almost notice it. The most noticeable difference in the sound field is the user's own voice. It sounds boomy and seems to come from inside the head. However, some users got used to this quite well. Eating and drinking sounds different and were considered undesirable by some users. These effects are caused by occluded ears and bone and tissue conduction (see Section 3.1.1). The problem is very difficult to overcome with the insert-type of headphones, which seal the ear canal tightly.

The sound color was perceived natural enough with the generic or individual equalization and the feel of direction and space was almost inseparable compared to the natural hearing situation, especially when there are visual clues (lack of visual clues deteriorate the frontal externalization to some degree). When listening to music with loudspeakers, some lack of the lowest and highest frequencies can be noticed but nothing dramatic (this might even be beneficial in normal speech communication situations). The lack of high frequencies could be corrected with proper equalization. The sound field was considered a little narrow when listening to loudspeakers through the headset. Driving a car and listening to a car stereo was experienced normal compared to the natural situation.

The noise of the microphones is hidden behind background noise in normal situations but can be noticed when it is very quiet. However, the test subjects reported that they got used to the noise very quickly and after a while they did not even notice it [36].

Ergonomics

The fitting is very critical in terms of user comfort. One of the test subjects had smaller ear canals, so he reported irritation of the ear after using the headset for a couple of hours.

Some critique also emerged about the noise and limitations caused by the wires of the headset. When the wires make contact with objects (for example the collar of the shirt or the zip of a jacket) they make undesired noise. They might also get stuck and limit the motion of the head. One way to overcome this problem would be to use very thin and flexible wires or even tape the wires to the body. Bluetooth or other wireless techniques could be also used to route the signals between the headset and the mixer, so no wires would be needed.

Talking to a mobile phone was also possible, but some inconveniences were recorded. The earphones fit quite tightly to the ear canal entrance, while the microphones remain about 1 cm out from that point (see Fig. 5.8), which could mean degradation of spatial perception and inconvenience of using a telephone. This could be overcome by using a smaller headset deeper inside the outer ear.

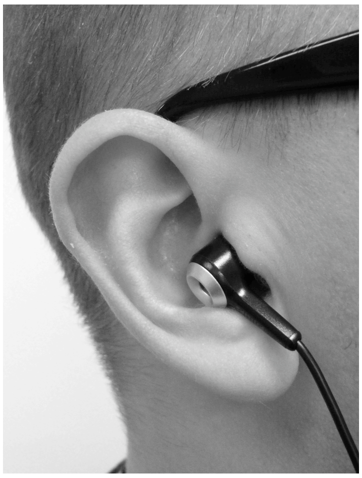

Figure 5.8: Headset inserted into the ear canal entrance.

Communication

The level of the user's own voice can be difficult to adjust at first and the user does not know if he/she is talking too loud or quiet. It is also possible that other people might not know that the subject who is wearing the ARA headset can still hear normally. This can cause inconvenient situations for example in grocery stores, when the cashier does not know whether the customer hears or not. This might be a problem also in a lecture hall if the lecturer assumes that the student is listening to music during the lecture.

All in all, the testees did not experience any problems in normal conversation situations. The sound quality was good and the test subjects quickly became used to their own voice also.

Other Notes

Loud and impact noises can cause some distortion to the pseudoacoustic representation. In normal life this is quite rare but can be noticed for example when closing the door of a car or listening to very loud music. One of the test subjects reported some distortion when listening to music above A-weighted SPL of 85 dB. The distortion may be caused by the microphone or the operation amplifiers of the circuit. The connectors of the wires could cause undesired scratchy noises when there were connector contact problems due to movement of the wires.

Some of the test subjects noticed noise caused by wind hitting the microphone. This might become more of a problem in strong wind or when moving fast, for example with a bicycle, but in normal circumstances it is acceptable.

The subject's own steps might sound a little boomy. When the subject runs or jumps, the headset also moves and makes additional sounds. If the exercise causes the subject to lose one's breath, the breathing starts to sound loud and unpleasant. This is quite similar to the effect of one's own boomy voice and is caused by bone and tissue conduction.

Feature Experimenting and Suggestions

It is possible to add external sounds to the pseudoacoustic representation and some of the users also experimented this with a portable audio device (MP3-player or suchlike). As in normal headphone listening, there was noticeable in-head localization but it can be even desirable when the subject wants to observe the environment at the same time. This way the music and the pseudoacoustic representation are kept separate and the subjects were pleased that they could hear the surrounding sound environment at the same time while listening to music. It is a property of the Philips headphone that low frequencies are a little bit pronounced, and this was reported by some users. The equalization is only done to the binaural microphone signals, not to the added external sounds.

If the goal is to embed the added sounds to the natural environment, then signal processing (for example using HRTFs) could be used. If the user would like to concentrate completely on the music or otherwise block the natural sound environment, it would be easy to add a noise canceling section to the mixer. It is also very important to be able to turn the music off quickly and easily during for example a conversation.

One of the desired features was the easy adjustment of the level of the pseudoacoustics.

In the prototype, the gain controls were only available with a small screwdriver, which was found to be inconvenient. However, the level should be controlled automatically to prevent hearing damage due to loud sounds in the environment or the user's own external sound sources (like an mp3-player).

One of the features the subjects suggested was interconnection with a mobile phone. The ARA headset and the mixer could be used as a hands-free system with a mobile phone. This would require proper connectors in the phone. In modern mobile phones there are also media players, so music could be played from the mobile phone to the headset.

In the design and construction of the ARA mixer, we did not concentrate very much in electronic design such as power consumption, the size of the circuit board or disturbance elimination. These were considered irrelevant for the study, but received some critique from the users as described earlier. If the headset and mixer combination was to be made commercial, these aspects would be critical in determining its success. However, the problems are quite easily overcome by proper design. The target was to make a better pseudoacoustic representation, and in that sense, the headset and mixer with equalization proved to perform very well both in the listening tests and in the field tests.

The test subjects reported that they would be willing to use the headset for longer periods of time if it was considered useful, and interesting ARA applications were available [36].

5.1.4 Discussion on the Listening and Usability Tests

The best results should be obtained when thorough in-ear measurements are conducted, and the mixer is carefully equalized for each user and the headset. The reason for this is that there are differences in the ear canal resonances and fittings of the earphone between users. However, it is not always possible to conduct proper measurements due to lack of time or facilities and hardware. In this case, it is possible to find equalization settings, which are generically suitable for most people. This can be done by collecting a large enough database of representable measurements from a set of users and equalize the mixer based on these findings. Even with four test subjects it was possible to find a good average equalization curve (Figure 5.4). Another way to overcome this problem would be to use adaptive equalization, which means that as the fitting of the earplug and the length of the ear canal varies between different users, the system would notice this and make an optimized equalization curve for the situation. This would, however, possibly require microphones also inside the closed ear canal and a sophisticated controlling system.

Due to a restricted number of filter sections, there is no possibility to find an exact filter, so the equalization is always a compromise. Therefore, tuning the parameters requires measurements and a few trial and fine-tuning rounds for optimal settings. This might also explain why some test subjects found the non-individual equalization better.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

As a conclusion, we found that with simple parametric magnitude equalization the overall sound quality of an ARA headset can be clearly improved. The usability studies and preliminary listening test results imply that a headset could be designed for everyday life usage and that this kind of generic mixer could be used to improve the sound quality of a headset. This way, the pseudoacoustic representation of the surrounding environment can be very close to the natural hearing experience. Of course, useful ARA applications are needed, since the pseudoacoustics as such does not bring added value to the user.

We also found that the headset could be used for acoustic head-tracking. Preliminary measurements imply that tweeters could be used as anchor sound sources, and the signal-to-noise ratio of the signals captured by the binaural microphones is sufficient. Six tweeters were installed to the walls of a room and pulse-type anchor signals were played from the tweeters and recorded with the binaural microphones of the Philips SHN2500 headphone. The signals were placed at frequencies from 18 kHz to 20 kHz, so they are almost non-audible for most people. The development of the positioning system is still under progress.

6.2 Future Work

The usability experiment was a preliminary study, and in the future we plan to further study the headsets with more extensive usability evaluations and listening tests with a larger group of people. Also more research for better non-individual equalization curves based on a larger group of subjects needs to be done. As the digital technology becomes faster, low latency digital signal processing and equalization can be possible. For example FPGAs (Field Programmable Gate Arrays) are a promising technology for these kinds of applications.

As long as analog technology is used, some improvements can be made to the ARA mixer. For example, the layout size could be optimized and made smaller, level adjustment could be made easier, the power consumption and protection against interference made better, and so on. Some features could also be added to the mixer easily. For example noise canceling (by shifting the phase of the binaural microphone signals before feeding them back to the headset) would be simple to implement. The headsets also require more research. Other active noise canceling headphones or bone conduction headphones could be tested and used.

In the future, a head-tracking algorithm will be finished and tested it in the setup we already built. The acoustic head-tracking could be compared to other tracking methods like video-based tracking or electromagnetic sensors. Also more research for useful ARA applications needs to be done. The platform is planned to be used for implementing and testing novel two-way applications. Also new head-tracking techniques should be studied, since outdoors mobility would be desirable. As wireless technology develops, some mobility and usability issues could be overcome. Especially the headset should be wireless and this could be implemented for example with Bluetooth -type of technology. The latest mobile phones already have Bluetooth, compass, WLAN, 3G, and other features, which could be useful for many ARA applications.

Bibliography

[1] G. Baratoff and S. Blanksteen. Tracking devices, http://www.hitl.washington.edu/scivw/eve/i.d.1.b.trackingdevices.html.

[2] D. R. Begault. *3-D sound for virtual reality and multimedia*. AP professional, 1994.

[3] E. H. Berger and R. W. Kieper. Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways. *J. of the Acoustical Society of America*, 114:1955–1967, October 2003.

[4] J. Borwick. *Loudspeaker and Headphone Handbook*. Focal Press, 2001.

[5] ITU Recommendation ITU-R BS.1284-1. General methods for the subjective assessment of sound quality.

[6] T. Caudell and D. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. *Proc. 25th Hawaii Int. Conf. on System Sciences*, II:659–669, January 1992.

[7] H. Dillon. Hearing aids. *Boomerang Press*, Sydney, 2001.

[8] A. Farina. Simultaneous measurements of impulse response and distortion with a swept-sine technique. *Proc. 108th AES Convention*, Paris, France, 2000.

[9] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho. Augmented reality audio for mobile and wearable appliances. *J. Audio Eng. Soc.*, 52(6):618–639, June 2004.

[10] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, H. Nironen, and S. Vesa. Techniques and applications of wearable augmented reality audio. *Proc. AES 114th Convention, Amsterdam*, The Netherlands, March 2003.

[11] P. Hiselius. Attenuation of earplugs - objective predictions compared to subjective reat measurements. *Acta Acustica united with Acustica*, 91:764–770, 2005.

[12] P. Hiselius. Acoustic modeling of earplugs. *Acta Acustica united with Acustica*, 92:135–138, 2006.

[13] M. Karjalainen. *Kommunikaatioakustiikka*. Helsinki University of Technology, 1999.

[14] M. Karjalainen, M. Tikander, and A. Härmä. Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources. *Proc. of IEEE Int. Conf. on Acoust., Speech, and Sig. Proc.*, May 2004.

[15] R. G. Keen. Simple, easy parametric and graphic eq's, plus peaks and notches, www.geofex.com/article_folders/eqs/paramet.htm, 1999.

[16] M. C. Killion. Earmold acoustics. *Seminars in Hearing*, pages 299–312, 2003.

[17] C. Kyriakakis. Fundamental and technological limitations of immersive audio systems. *Proc. IEEE*, 86:941–951, May 1998.

[18] C. Kyriakakis, P. Tsakalides, and T. Holman. Surrounded by sound. *IEEE Signal Process. Mag.*, 16:55–66, January 1999.

[19] C. Kyriakis and T. Holman. Video-based head tracking for improvements in multichannel loudspeaker audio. *Proc. 105th AES Convention*, September 1998.

[20] T. Lokki, H. Nironen, S. Vesa, L. Savioja, A. Härmä, and M. Karjalainen. Application scenarios of wearable and mobile augmented reality audio. *Proc. AES 116th Convention*, Berlin, Germany, May 2004.

[21] K. Mayer, H. L. Applewhite, and F. A. Biocca. A survey of position-trackers. *Presence: Teleoperators and virtual environments*, 1(2):173–200, 1992.

[22] S. Mehrgardt and V. Mellert. Transformation characteristics of the external human ear. *J. of the Acoustical Society of America*, 61:1567–1576, April 1976.

[23] H. Møller, C. B. Jensen, D. Hammershøi, and M. F. Sörensen. Design criteria for headphones. *J. Audio Eng. Soc.*, 43(4):218–232, April 1995.

[24] New York State Department of Health. Sectional view of an ear, http://www.health .state.ny.us/nysdoh/antibiotic/4815.htm.

[25] B&K HATS 4128 Product Picture. http://www.bkhome.com/ip_template.asp?upid=9985pf=1.

[26] C. Poldy. Headphone fundamentals. *Tutorial presented in the AES 120th Convention*, Paris, http://www.aes.org/tutorials/, May 2006.

[27] V. Riikonen, M. Tikander, and M. Karjalainen. An augmented reality audio mixer and equalizer. *Accepted to be presented in the 124th AES Convention*, Amsterdam, May 2008.

[28] FuzzMeasure software. http://www.supermegaultragroovy.com/products/fuzzmeasure/.

[29] Pure Data software. http://puredata.info/.

[30] H. Y. Sung, J. Kim, and S. Jang. A method for objective sound quality evaluation of headphones. *AES 32nd INTERNATIONAL CONFERENCE*, pages 83–87, September 2007.

[31] Philips Product Support. Philips SHN2500 product picture, http://www.dealersupport .se /imagebank/images/shn2500.jpg.

[32] M. Tikander. Sound quality of an augmented reality audio headset. *Proc. of the 8th Int. Conference on Digital Audio Effects (DAFx'05)*, pages 178–181, Madrid, Spain, September 2005.

[33] M. Tikander. Modeling the attenuation of a loosely-fit insert headphone for augmented reality audio. *AES 30th INTERNATIONAL CONFERENCE*, Saariselkä, Finland, March 2007.

[34] M. Tikander, A. Härmä, and M. Karjalainen. Binaural positioning system for wearable augmented reality audio. *Proc. of the IEEE Workshop on Appl. of Sig. Proc. to Audio and Acoust. (WASPAA'03)*, pages 153–156, October 2003.

[35] M. Tikander, A. Härmä, and M. Karjalainen. Acoustic positioning and head tracking based on binaural signals. *Proc. 116th AES Convention*, page 10, May 2004.

[36] M. Tikander, M. Karjalainen, and V. Riikonen. An augmented reality audio headset. *Submitted to Proc. of the 11th Int. Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, September 2008.

[37] S. E. Voss and J. B. Allen. Measurement of acoustic impedance and reflectance in the human ear canal. *J. of the Acoustical Society of America*, 95:372–384, September 1993.

Appendix A

Controlling the Equalization Parameters of the ARA Mixer

As a default equalization setting, there is an average curve taken from four persons. It is based on the curve shown in Figure 5.4.

If there is a need to change the settings, the cover of the box must be opened and these instructions followed. For the operation, an electronics screwdriver, two wires (male mini-plugs in both ends) and some software to measure the frequency response are needed. For example, FuzzMeasure software [28] can be used to measure the frequency response of the ARA mixer.

First the power has to be turned on and one wire connected to the "Mic" connector of the mixer and the other end to the output of the computer's sound card. Then the other wire has to be connected to the "Phones" connector of the mixer and the other end to the input of the sound card. Now there is a signal route that allows the user to measure the frequency response of the ARA mixer from the microphone connector to the headphone connector. This should give a flat frequency response when the equalization is turned off (middle position of the three position slide switch). Small levels have to be used so that the mixer does not break.

After the setup is ready, the adjustment of the equalization parameters can begin. The equalization can be turned on from the slide switch. The equalization properties can be changed from the trimmers with a small electronics screwdriver (the trimmers for the right channel are 1-7 in Figure A.1). The frequency trimmers (4, 5, 10, 11) do not have stoppers, so they rotate freely.

Adjustment can be done to the bass reproduction and to the two parametric equalization sections (for both channels naturally). The first one is supposed to be used as a peak and the second one as a notch, but they can be both used as peaks or notches if wanted.

The gain trimmers (2, 7, 8, 13) determine the height of the peak. The Q-value sets the steepness and the width of the peak (the higher the value, the narrower and steeper the peak is). The center frequency trimmer determines the place of the peak or notch. The first peak can be moved in the range of 700-3200 Hz and the second notch between 1800-8500 Hz. The cut-off frequency of the bass control (-3 dB) can be moved from 6 to 700 Hz. For more information about the equalization and the rationale behind it, see [27].

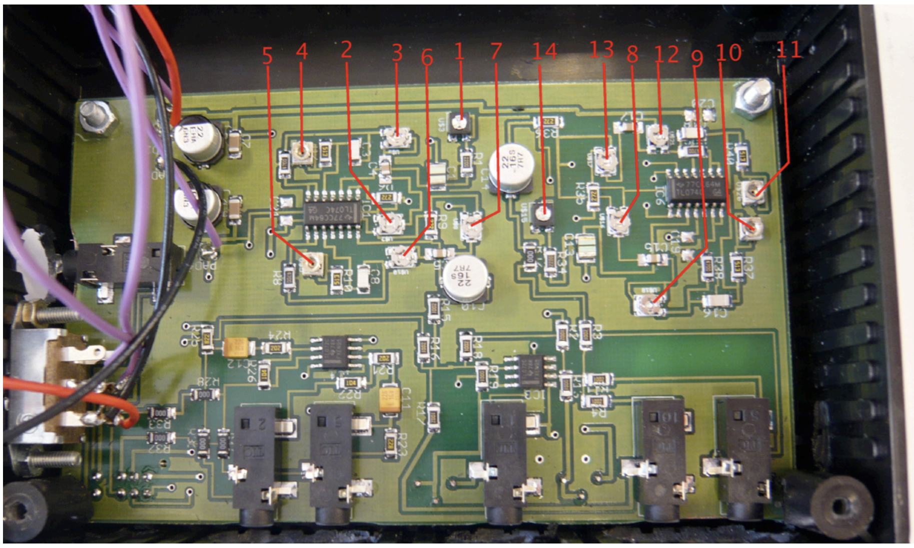

Figure A.1: The ARA mixer equalization controls

1. Bass control (R): turn clockwise to add bass
2. Gain of the first peak (R): turn clockwise to make a peak
3. Q-value of the first peak (R): turn anti-clockwise to add Q
4. Frequency of the first peak (R): turn clockwise to increase the center frequency
5. Frequency of the second notch (R): turn anti-clockwise to increase the center frequency
6. Q-value of the second notch (R): turn clockwise to add Q
7. Gain of the second notch (R): turn clockwise to make a notch
8. Gain of the first peak (L): turn clockwise to make a peak
9. Q-value of the first peak (L): turn anti-clockwise to add Q

10. Frequency of the first peak (L): turn anti-clockwise to increase the center frequency

11. Frequency of the second notch (L): turn clockwise to increase the center frequency

12. Q-value of the second notch (L): turn clockwise to add Q

13. Gain of the second notch (L): turn clockwise to make a notch

14. Bass control (L): turn clockwise to add bass

Appendix B

Summary of the Usability Diaries

Table B.1: *Diaries: The higher the attribute, the better the quality. ++++ = all testees found an attribute working, +++ - = all but one testee found the attribute working,... - - - - = all testees found the attribute not working*

Attribute	Observations
++++	
Directional hearing	No difference to natural listening
Spatial impression	Practically no difference to natural listening
Speech intelligibility	Once accustomed to ones own voice, there were no problems having conversations
Walking	Initially own steps sounded boomy but after a while walking was felt natural
+++-	
Sound quality	"*Higher frequencies missing*", "*Timbre very close to natural*"
Wind	Moderate wind ok, strong wind overloaded the microphones resulting in distortion
Inherent noise	Audible only in very quiet environments
Ear ache	One of the testees with small ears had problems with ear ache
Watching TV	Almost natural experience
Concert	Loud music overloads the microphone or the electronics resulting in unpleasant distortion
Listenig to music	High frequencies attenuated, "*Sound stage doesn't open*"
++ - -	
Distortion (impulses)	With loud bangs, the mixer was overloaded and the sound was distorted. Some testees got used to it while others did not.
+ - - -	
Distortion (cont.)	Continuous overload distortion due to loud environment was irritating
Eating	"*The chewing noise was very annoying*"
- - - -	
Talking to a mobile phone	The headset extended too far outside the ear
Running/jumping	Hard steps and jumping sound annoyingly boomy
Strong wind	Strong wind and running can make unpleasant distortion
Size of the mixer box	The prototype ARA mixer box was fairly large

Appendix C

Circuit Diagram and Components of ARA Mixer

Figure C.1 shows the complete circuit diagram for one channel of the ARA mixer. The component values are in the tables below.

Table C.1: Resistor values of the ARA mixer circuit (see Figure C.1).

Component	Value	Component	Value
R_1:	8.2 kΩ	R_{15}:	470 Ω
R_2:	2 kΩ	R_{16}:	51 kΩ
R_3:	100 kΩ	R_{18}:	10 kΩ
R_4:	220 Ω	R_{19}:	10 kΩ
R_6:	2.7 kΩ	R_{20}:	10 kΩ
R_9:	2.7 kΩ	R_{21}:	10 kΩ
R_{11}:	470 Ω	R_{22}:	100 Ω
R_{12}:	51 kΩ		

Table C.2: Capacitor values of the ARA mixer circuit (see Figure C.1).

Component	Value
C_1:	22 μF
C_2:	1 μF
C_3:	22 μF
C_4:	33 nF
C_5:	3.3 nF
C_6:	15 nF
C_7:	1 nF

Table C.3: Trimmer resistor values of the ARA mixer circuit (see Figure C.1).

Component	Value
R_5:	20 kΩ
R_7:	10 kΩ
R_8:	10 kΩ
R_{10}:	10 kΩ
R_{13}:	1 MΩ
R_{14}:	10 kΩ
R_{17}:	1 MΩ
R_{23}:	10 kΩ

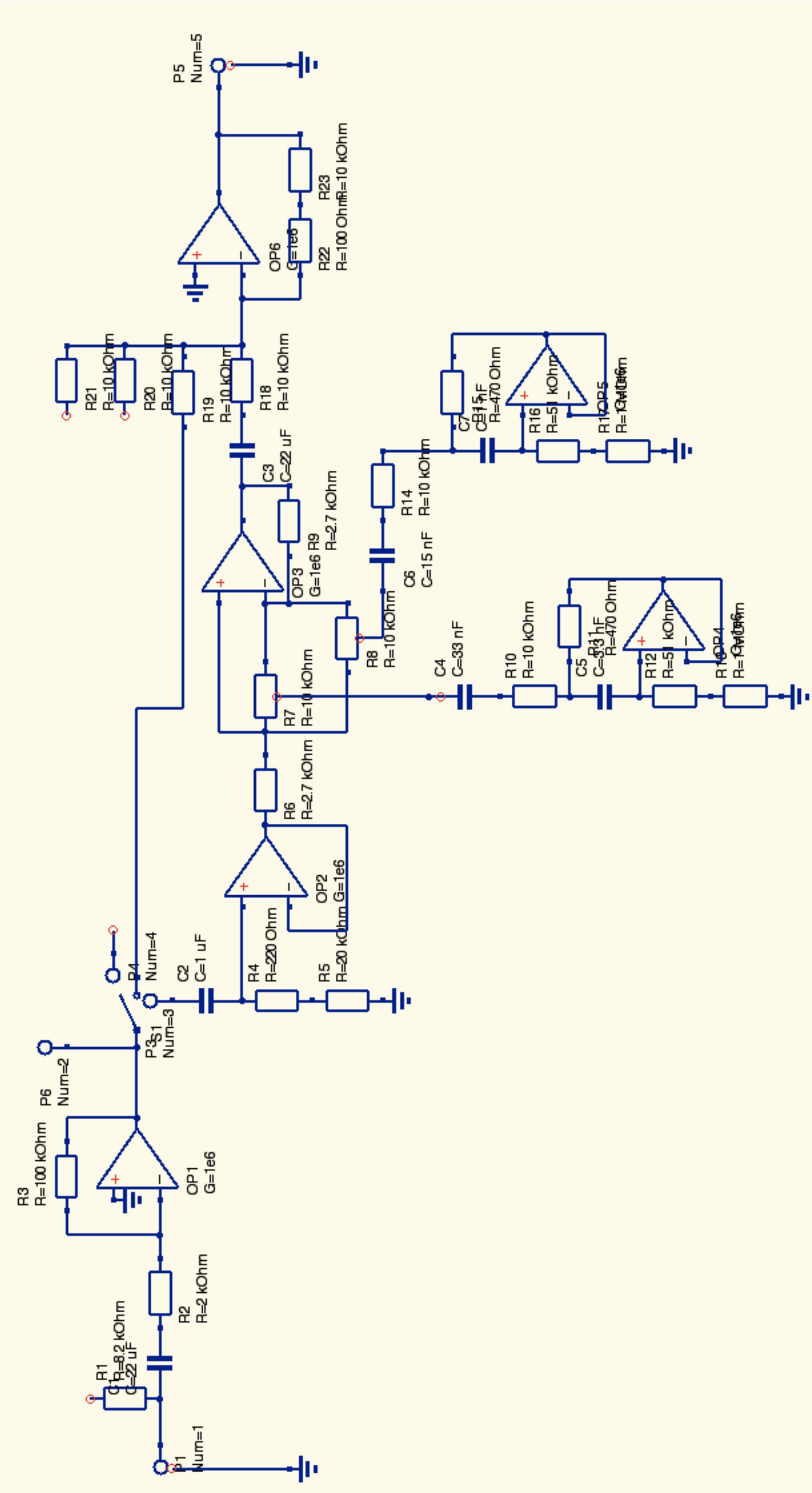

Figure C.1: The complete circuit for one channel of the ARA mixer.