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

Jussi Rämö

Evaluation of an Augmented Reality Audio Headset and Mixer

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

Espoo, May 5, 2009

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

HELSINKI UNIVERSITY OF TECHNOLOGY	ABSTRACT OF THE MASTER'S THESIS

Author:	Jussi Rämö		
Name of the thesis:	Evaluation of an Augmented Reality Audio Headset and Mixer		
Date:	May 5, 2009	**Number of pages:**	73
Faculty:	Faculty of Electronics, Communications and Automation		
Professorship:	S-89		
Supervisor:	Prof. Matti Karjalainen		
Instructor:	Miikka Tikander, M.Sc.		

Augmented Reality Audio (ARA) is a concept that is defined as a real-time combination of real and virtual auditory worlds, that is, the everyday sound surroundings can be extended with virtual sounds. The hardware used in this study for augmented reality audio consists of a pair of headphones and a controlling unit, called an ARA mixer. The ARA headphones are composed of binaural earphone elements with integrated microphones. The ARA mixer provides all the connections and signal processing electronics needed in ARA applications.

The basic operating principle of the ARA headset is that the binaural microphones should relay the sound signals unaltered to the earphones in order to create an accurate copy of the surrounding sound environment. Unfortunately, the ARA headset creates some alterations to the copied representation of the real sound environment. Because of these alterations, the ARA mixer is needed to equalize the headphones. Furthermore, the ARA mixer enables the addition of virtual sound objects. Virtual sound objects can be embedded into the real environment in a way that the user can distinguish them from the real sound environment or in a way that the user cannot tell the difference between the real and virtual sounds.

The aim of this thesis is to perform full-scale laboratory measurements and an usability evaluation of the ARA hardware. The objective is to collect technical data about the hardware and to gather knowledge concerning how users perceive the usability of the ARA headset in everyday-life situations. With the gathered information it is possible to further improve the usability and sound quality of the ARA hardware.

Keywords: Augmented Reality Audio (ARA), ARA headset, In-ear headphone measurements, In-ear headphone usability

TEKNILLINEN KORKEAKOULU — DIPLOMITYÖN TIIVISTELMÄ

Tekijä:	Jussi Rämö		
Työn nimi:	Lisätyn audiotodellisuuden laitteiston evaluointi		
Päivämäärä:	5.5.2009	**Sivuja:**	73
Tiedekunta:	Elektroniikan, tietoliikenteen ja automaation tiedekunta		
Professuuri:	S-89		
Työn valvoja:	Prof. Matti Karjalainen		
Työn ohjaaja:	DI Miikka Tikander		

Lisätty Audiotodellisuus (LAT) on käsite, joka on määritelty todellisen ja virtuaalisen maailman reaaliaikaisena yhdistelmänä. Täten jokapäiväiväiseen äänimaailmaan voidaan lisätä virtuaalisia ääniobjekteja. Lisätyn audiotodellisuuden laitteisto, jota tutkitaan tässä työssä, koostuu kuulokeparista sekä kontrolliyksiköstä, nimeltään LAT-mikseri. LAT-kuulokkeet koostuvat binauraalisista kuuloke-elementeistä sekä sisäänrakennetuista mikrofoneista. LAT-mikserissä on kaikki LAT-sovellusten tarvitsemat liittimet sekä signaalinkäsittelyelektroniikka.

LAT-kuulokkeiden toimintaperiaate perustuu siihen, että binauraalisten mikrofonien tulisi välittää äänisignaalit muuttumattomana kuuloke-elementeille, jotta todellinen äänimaailma saataisiin kopioitua muuttumattomana. Valitettavasti LAT-kuulokkeet aiheuttavat muutoksia kopioituun äänimaailmaan. Näiden muutoksien takia tarvitaan LAT-mikseriä ekvalisoimaan kuulokkeita. LAT-mikseri mahdollistaa myös virtuaalisten ääniobjektien lisäämisen. Virtuaaliset ääniobjektit voidaan lisätä todellisen äänimaailmaan siten, että käyttäjä voi erottaa ne todellisesta äänimaailmasta tai siten, että käyttäjä ei erota virtuaalisia ja todellisia äänilähteitä toisistaan.

Tämän diplomityön tavoitteena on mitata LAT-laitteiston suorituskykyä erilaisten laboratoriomittausten avulla sekä suorittaa käyttäjäkoe. Mittausten ja käyttäjäkokeen avulla pyritään selvittämään LAT-laitteiston tekniset tiedot sekä ymmärtämään miten käyttäjät kokevat LAT-laitteiston käytettävyyden jokapäiväisessä elämässä. Kerätyn informaation avulla on mahdollista kehittää LAT-laitteiston käytettävyyttä sekä äänenlaatua.

Avainsanat: Lisätty Audiotodellisuus (LAT), LAT-kuulokkeet, In-ear kuulokeiden mittaukset, In-ear kuulokkeiden käytettävyys

Acknowledgements

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

I want to thank Professor Matti Karjalainen for excellent guidance with this thesis and for the possibility to work in this interesting project. Matti's feedback during the writing process have been very helpful.

I wish to thank M.Sc. Miikka Tikander for all the assistance and knowledge he has provided during my thesis writing.

I would also like to thank my co-workers in the department, especially Marko Hiipakka for all the collaboration, and all of them who participated in the usability evaluation. Furthermore, I would like to thank my fellow student Markus Pitkäranta for all the support and good times during my studies.

My gratitude also goes to Julia Turku, Riitta Väänänen, Matti Hämäläinen and Johan Kildal from Nokia Research Center, and to Lauri Savioja, Tapio Lokki and Mikko Peltola from the Department of Media Technology.

Finally, I would like to thank my friends and family, especially my parents Arto Rämö and Heli Salmi for all the support during my studies.

Otaniemi, May 5, 2009

Jussi Rämö

Contents

Abbreviations

ITD	Interaural Time Difference
ILD	Interaural Level Difference
HRTF	Head Related Transfer Function
AR	Augmented Reality
ARA	Augmented Reality Audio
MARA	Mobile Augmented Reality Audio
WARA	Wearable Augmented Reality Audio
KAMARA	Killer Applications in Mobile Augmented Reality Audio
ANC	Active Noise Canceling
MP3	MPEG 1 Audio Layer 3
GPS	Global Positioning System
WLAN	Wireless Local Area Network
3-D	Three-Dimensional
DOF	Degrees Of Freedom
LED	Light Emitting Diode
VoIP	Voice Over IP
IP	Internet Protocol
ADECS	Adjustable Ear Canal Simulator
THD	Total Harmonic Distortion
RMS	Root Mean Square
GSM	Global System for Mobile Communications
EMC	Electromagnetic Compatibility
MUX	Multiplexer
MOS	Mean Opinion Score

Chapter 1

Introduction

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

Since then, the project has been ongoing both at the Nokia Research Center and Helsinki University of Technology. Since the beginning of 2005, two doctoral students have participated in the KAMARA Project with academic funding. The collaboration between NRC and the two laboratories at Helsinki University of Technology continued from autumn 2007, called KAMARA+. The KAMARA+ Project resulted in two thesis, one for each laboratories, which were completed in the spring of 2008.

KAMARA++ Project was started right after KAMARA+, in the spring of 2008. KAMARA++ is composed of two tasks; evaluation of KAMARA platform (Department of Signal Processing and Acoustics), and application and KAMARA platform development for outdoor use (Department of Media Technology). This thesis concentrates on the user-related acoustics, including sound quality and usability issues. At Department of Media Technology the concentration is on general platform architecture and implementation of the outdoor application.

Augmented Reality (AR) is defined as a real-time combination of real and virtual worlds [15]. In Augmented Reality Audio (ARA) the everyday sound environment can be extended with virtual sound objects. In order to embed the virtual sound objects into the natural sound environment special equipment are needed. The specific ARA hardware has been developed in [30] [31]. The ARA hardware contains a hear-through ARA headset and an ARA mixer, which provides all the connectors and signal processing electronics needed in

ARA applications.

Preliminary analysis and testing of the ARA hardware, including a small-scale usability test, has been conducted in [37]. The goal of this thesis was to perform detailed laboratory measurements and a full-scale evaluation of use of the mixer and headset. A number of laboratory measurements were conducted in order to determine the technical specifications of the ARA headset and mixer. The goal of the usability test was to find out what works well and what are the problems in lengthened use of the ARA hardware.

Chapter 2 of this thesis introduces the background related to the basics of hearing. Chapter 3 describes the ARA technology and head-tracking in general. The documentation of the laboratory measurements is presented in Chapter 4. Chapter 5 presents the usability test of an augmented reality audio headset and Chapter 6 introduces the conclusions of this research and suggests what should be done in the future.

Chapter 2

Basics of Hearing

Hearing is a very important sense to human beings along with the vision. The ability to hear and see helps us to be aware of our environment. Hearing also makes communication between people possible, which is something unique compared to all other animals. The human ability to communicate through speech is also reflected in the structure of the ear, discussed in Section 2.1. The ear is divided into three parts; outer ear, middle ear and inner ear.

Hearing Event

The pinna (or ear flap) funnels sound waves into the auditory canal (or ear canal) where the sound wave resonates according to the acoustic properties of one's auditory canal. When the sound wave gets to the ear drum it makes the ear drum to vibrate and transfers the sound information to the ossicles in the middle ear. The ossicles pass the vibration onwards to the oval window, which connects the middle ear and the inner ear. At the inner ear side of the oval window is the cochlea that is filled with fluid, thus the vibrations of the oval window create movement in the fluid. The basilar membrane, which is located in the cochlea (within the fluid), has many hair cells, which react to the liquid's movement. As a result, the hair cells send impulses to the auditory nerve and this way to the brain where the actual interpretation of sound happens. See Figure 2.1 for the anatomy of the ear.

To conclude, the pressure changes in the environment are funneled by the pinna, transported through the auditory canal, impedance matched between air and fluid by middle ear, then altered to movement of the fluid in the cochlea and finally the impulses of the hair cells are transported to the brains via the auditory nerve. Thus, the hearing system converts sound energy to mechanical energy and further to nerve impulses.

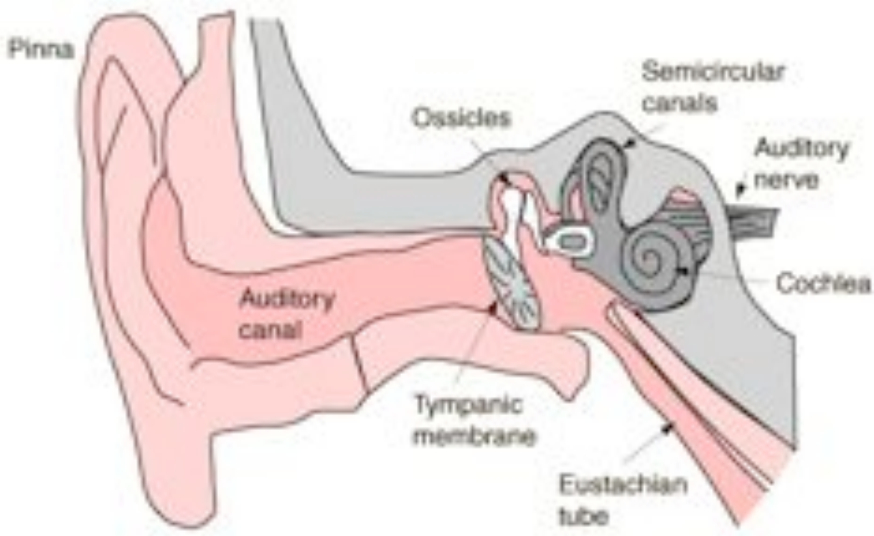

Figure 2.1: Anatomy of human ear [2]

2.1 Structure of the Outer Ear

In this study the main interest, concerning the anatomy of hearing, is within the outer ear. Outer ear consists of the pinna, the auditory canal and the ear drum (or tympanic membrane). The only function of the outer ear is to supply and filter the upcoming sound to the middle ear. Pinna is the visible part of the ear, shown in Figure 2.1.

The main task of the pinna is to collect and filter sounds. It amplifies and directs the sounds into the auditory canal. The filtering enhances the frequencies in the frequency area where human speech is located and it also gives cues for the direction of the sound. Without pinna it would be very hard to know if a sound is coming in front of or from behind us. The pinna has the greatest impact at high frequencies. It emphasizes the sounds coming in front and attenuates the sounds coming from behind.

The auditory canal (see Figure 2.1) is a tube that provides a passageway from the pinna to the ear drum. The average length and diameter of an auditory canal are approximately 22.5 mm and 7.5 mm, respectively [19]. The auditory canal can be considered to be a rigid tube, which ends at a rather hard ear drum with a frequency-dependent acoustical impedance. It acts as a quarter-wavelength resonator, where one end is being open and one is being closed. The first resonance of the auditory canal emphasizes the important regions for speech intelligibility. Thus, the auditory canal enhances the speech-specific frequencies as well, in addition to the pinna's enhancement.

The quater-wavelength resonance can be calculated with

$$f_0 = \frac{c}{4l}, \tag{2.1}$$

where c is the speed of sound and l is the length of the ear canal. Using the above-mentioned average length of the auditory canal, room temperature and Equation (2.1), the first ear canal resonance occurs approximately at 3.8 kHz. Sound velocity in room temperature (21 °C) is about 344 m/s, but if the resonance is calculated near body temperature (36 °C), where the sound velocity is about 353 m/s, the resonance moves to 3.9 kHz. Attached mass effect [20] is also present, which means that the effective length of the auditory canal is slightly longer than the physical length. This should be taken into account as well. When the length of the auditory canal increases, its resonance frequency decreases. Normally the first quarter-wavelength resonance is around 2-4 kHz.

The ear drum transforms the pressure changes arriving via the auditory canal to mechanical vibration of the ossicles.

2.2 Binaural Hearing

Humans, as well as other vertebrates, normally have two ears. Hearing with two ears is called binaural hearing. Binaural hearing brings a lot of benefits compared to monaural hearing (hearing with one ear only). Our brains can elaborately interpret impulses sent by both ears, thus helping us to be more aware of our acoustic surroundings. The fact that our ears are located at different sides of our head is not a coincidence; in fact, it is an income of evolution. It is a very important feature because it allows us to locate different sound sources.

2.2.1 Localization and Binaural Cues

Binaural cues are based on the above-mentioned fact that we have two ears and therefore two different auditory inputs. If the sound source is not exactly in front of us there will be time and level differences between the right and left ear. These cues help us to locate sound sources. Without binaural cues it is very difficult and the sound source is often perceived inside one's head. This is called lateralization (see Section 2.2.4) and it is a common phenomenon, for example, in headphone listening.

Localization has been extensively investigated and it has been widely accepted that localization depends upon binaural cues; ITD and ILD [11]. For complete localization (in free-field with one sound source), we need three parameters concerning the location of the sound source, these are azimuth (i.e., horizontal angle), elevation (i.e., vertical angle) and distance. With these parameters we can use spherical polar coordinates and place the sound source anywhere in space.

Azimuth is much more accurate to evaluate than elevation, partly because ITD and ILD provide binaural cues mainly in the horizontal plane. The estimation of distance, in anechoic space, is primarily based on the sound intensity instead of the actual distance, whereas in normal echoic listening situations the estimation is constituted with the help of the echoes and reverberations of the listening environment [19]. The distance estimation is quite inaccurate as well. It helps if the sound source is previously know by the listener, because then he/she has an intuition what the sound should sound like and with what intensity (for example, an incoming car). In fact, learning and adaptation are important parts of one's individual localization.

2.2.2 Interaural Time Difference - ITD

Interaural Time Difference (abbreviated ITD) is an important binaural cue, as mentioned before. Only when the sound source is exactly in front of a subject the distance to both ears is the same. If the sound source is located somewhere else, for example, right side of the listener (as in Figure 2.2), the wavefront must travel different distances to right and left ear. This means that the wavefront arrives to each ear at slightly different time, in this case, the right ear is reached first. The maximum time difference is a bit under 700 μs when the sound is coming directly on either side of the head [19]. The elevation also affects the ITD.

2.2.3 Interaural Level Difference - ILD

The other important binaural clue is Interaural Level Difference (abbreviated ILD). ILD measures the sound pressure level difference between the ears. Different sound pressure levels arise because the head shadows the incoming wavefront (see Figure 2.2). Therefore, the ear closer to the wavefront receives a higher sound pressure level than the shadowed ear. ILD is highly frequency-dependent because the head shadows high frequencies more than low frequencies. When the wavelength becomes much longer than the diameter of the head, the shadowing effect diminishes. Then the head is so small compared to the wavelength that it does not create big enough obstacle in front of the wavefront to create a shadowing effect. The sound waves just diffract and bend around the head. The pinna has also some influence to the ILD. It boosts the sound pressure level of the ear closest to the sound source and increases the shadowing effect at the other ear.

Rayleigh's duplex theory asserts that ILD and ITD are complementary. At the frequencies higher than 1500 Hz there is less ITD information and the ILD dominates. Below 1500 Hz the situation is the other way around [7].

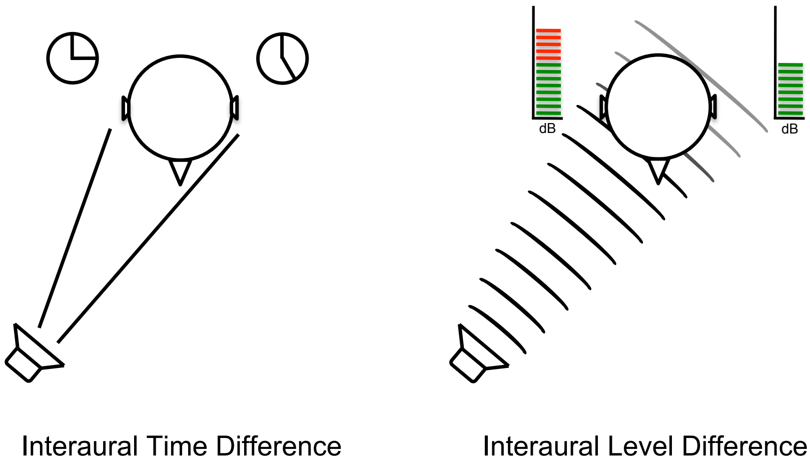

Figure 2.2: ITD and ILD are important binaural cues that help us to locate sound sources.

2.2.4 Head Related Transfer Function - HRTF

If ITD and ILD are measured with a symmetrical artificial head, the observation is that the elevation does not affect either one. Still, it is clear that we can perceive elevation of the sound source as well, at least with some accuracy [19]. That means that there must be some extra cues in addition to ITD and ILD. These (monaural) cues come from the human anatomy. The antisymmetry of the head, pinna, shoulders and upper torso affect the sound waves, e.g., by filtering, colorizing and shadowing them. Depending on the angle of the incoming sound wave they also create different kind of reflections. Thus, evaluation of the elevation is based on the antisymmetry of the head and pinna as well as the reflections coming from the shoulders. All these effects together are depicted by so called Head Related Transfer Function (abbreviated HRTF).

Despite the name it considers the shoulders and upper torso as well. Head Related Transfer Functions are, intuitively, individual. Each person has slightly different anatomy of the upper body and therefore slightly different HRTFs as well. For example, the pinna is a complicated resonator system with all its individual arcs and notches.

HRTF is the transfer function from the sound source to the entrance of the auditory canal in free-field. The auditory canal and ear drum do not add anything to spatial and directional hearing, thus it is sufficient to measure HRTFs from the entrance of the auditory canal. Since the HRTFs are different from person to person, the best way would be to measure them individually, yet for practical reasons (time and money) averaged HRTFs from a set of subjects are often used. Another way is to use a dummy head (i.e., head and torso simulator), which has microphones planted in its ears. Dummy heads are based on average adult anthropometric data.

HRTF filters can be used to filter stereo or mono recordings and that way create spatial and directional effects to the perceived sound, hence creating virtual sound sources.

Lateralization

When listening with headphones the voices tend to localize inside the head of a listener. The majority of recordings out there is meant to be reproduced with loudspeakers. When listening this kind of material with headphones, it excludes the transfer function of the listening room, head and outer ear, in other words it excludes the listener's HRTFs. If the headphone signal is filtered with the individual HRTFs of the listener (or approximating that with averaged HRTFs), the voices seem to come out of the head nearer to a speaker-like reproduction.

Chapter 3

Overview of ARA Technology

3.1 The Concept of Augmented Reality Audio - ARA

Augmented Reality (AR) is defined as a real-time combination of real and virtual worlds [15]. Probably the most intuitive implementation of Augmented Reality is a visual see-through display, which shows the real world with extended virtual content (see Figure 3.1). The same concept is used in Virtual Reality Audio (abbreviated ARA) [15].

Figure 3.1: Visual augmented reality. Real-world vision extended with virtual information.

In Augmented Reality Audio the everyday sound surroundings can be extended with virtual sounds, such as telephone conversation or audio calendar [30]. It is implemented with hear-through headphones called ARA headset. The ARA headset is composed of binaural earphone elements with integrated microphones (the ARA hardware is introduced in Section 3.3). The basic idea of the ARA headset is that the microphones should relay the sound signals unaltered with minimal latency (<< 1 ms) to the earphones in order to create a copy of the surrounding sound environment. The copy that has gone through ARA headset is called pseudoacoustic environment [15].

The objective is to make the ARA Headset as acoustically transparent as possible. Unfortunately, that requires some work. The ARA headset creates some alterations to the pseudoacoustic representation of the real environment (mainly due to the headphone acoustics discussed in Section 3.2). Because of these colorations in the pseudoacoustic representation some equalization is needed. For this and in order to add virtual sound objects to the pseudoacoustic sound field a special device is needed. This device is called ARA mixer [30] (see Section 3.3).

Virtual sound objects can be embedded into the pseudoacoustic representation so that the user can distinguish them from the real sound environment or so that the user cannot tell which of the sounds are virtual and which are real. These virtual sounds can be almost anything from music to active noise cancelling or from teleconference to acoustic post-it stickers (see Section 3.5 for further possible application scenarios).

The basic system diagram for ARA applications is shown in Figure 3.2. The signals that user hears through an ARA headset are composed of pseudoacoustic signals captured by the binaural microphones and virtual input signals embedded to the pseudoacoustic environment in a desired way. The embedding happens in the ARA mixer. The preprocessing block can be used to send user's microphone signals to a distant user.

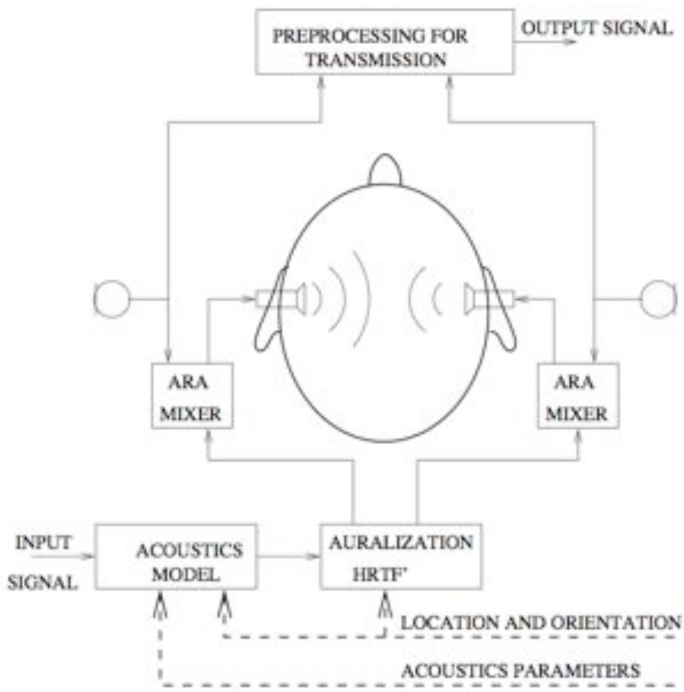

Figure 3.2: Generic system diagram for ARA applications [15]

MARA – Mobile Augmented Reality Audio and WARA – Wearable Augmented Reality Audio are complements to the concept of ARA. MARA simply means that the ARA is implemented so that is mobile, like the headphone implementation concerned in this thesis. The alternative implementation would be, for example, a fixed position loudspeaker setup in a listening room. MARA and WARA can mean the the same thing. WARA is a little stricter concept than MARA because the ARA hardware must be so compact that you can wear it (in everyday life). The headphone implementation covered in this thesis can be classified to be both MARA and WARA [15][16].

In order to create realistic augmented virtual services (i.e., position-aware applications) one key factor is the ability to track user's location, orientation and head position [36]. This can be done with head- and motion-trackers. There are many different approaches to these kind of head-tracking and positioning implementations (see Section 3.4).

3.2 Headphone Acoustics

Headphones are nowadays very common in everyday life. They are used for music listening, hands-free sets and hearing aid devices to name a few. One excellent feature in the ARA mixer is that it can combine a number of devices, which normally requires different headsets, into one ARA headset. There are many different types of headphones, which all have their pros and cons depending on the purpose of their use. Different applications also set different requirements for headphone design (such as size, quality and cosmetic properties).

One way to categorize headphones is to separate them in circumaural, supra-aural and in-ear headphones. Circumaural headphones are complitely around the ear (like a cup), while supra-aural headphones are placed on the ear flap. In this categorization the in-ear headphones includes both the "actual" in-ear headphones, that has a silicone tip which is put into the ear canal, and the more conventional earphones that are put in-ear, but which fit loosely in the concha and do not reach into the ear canal (see Figure 3.3). In this thesis the term "in-ear headphones" is used to mean the former type of headphones, which is actually put into the ear canal.

There has been quite much research done in the area of circum- and supra-aural headphones in contrary to in-ear headphones, although, in-ear headphones are very popular in this day and age. The ARA headset of this study is implemented with in-ear type headphones (see Section 3.3.1) so extensive measurements have been conducted while designing the ARA hardware [30].

Figure 3.3: Conventional in-ear headphones [34].

3.2.1 Normal vs. Headphone Listening

In normal listening situation (with open ears), e.g., when listening to music with loudspeakers or when walking in downtown listening the tumult going on there, the sound waves are modified by the listener's body, head and outer ear, as mentioned before. If these sounds are recorded normally with a microphone and then played through headphones, all these modifications caused by the body of the listener are lost. This must be taken into consideration when designing headphones. This is quite a challenging task to realize because every ear and body is more or less different (even the ears of one person differ from each other). Different headphone types also affect differently in the acoustics of the outer ear.

Ear Canal Resonances

Normally the ear canal is open and acts as a quarter-wave resonator as mentioned in Section 2.1. The effect of an open ear canal is shown in Figure 3.4. The curve represents the difference between the frequency response of a Genelec 8030A active monitor inside the ear canal (measured approximately 1 cm deep) and in free-field recorded with Knowles FG-3329 miniature microphone. As can be seen, the first two peaks (resonances) in Figure 3.4 appear at frequencies 3 kHz and 8 kHz. When an earplug (in-ear headphone) is inserted into the ear canal the acoustic properties of the ear canal changes. The earplug can go rather deep in the ear canal (approximately 5 mm), which means that the length of the ear canal becomes shorter and most of all it blocks the ear canal entrance making it a closed tube. A closed tube acts as a half-wave resonator and the first resonance occurs at

$$f_0 = \frac{c}{2l}, \tag{3.1}$$

where c is the speed of sound and l is the length of the ear canal. The closed ear canal does not only create a new half-wave resonance but it also cancels the quarter-wave resonance, created by the open ear canal, which people are used to. The half-wave resonance occurs normally in the range of 5 - 10 kHz depending the type of headphones and the structure of user's ears.

These resonance behaviors need to be taken into account when designing headphones (especially in-ear headphones). Basically headphones need to cancel the half-wave resonance and create the quarter-wave resonance in order to sound natural. Therefore, unlike with loudspeakers, the ideal frequency response of headphones is not flat. An ideal frequency response in "combined field" for headphones according to [25] is shown in Figure 3.5. There are two main ways to calibrate headphones, in free-field or in diffuse-field. Diffuse-field measurement takes all the reflections into account, which happens, e.g., in normal loudspeaker listening situations. When the distance from the loudspeaker is greater than the reverberation radius the sound field carries more reflected sound than direct sound. The combined field used in Figure 3.5 is an average of the free- and diffuse-field measurements.

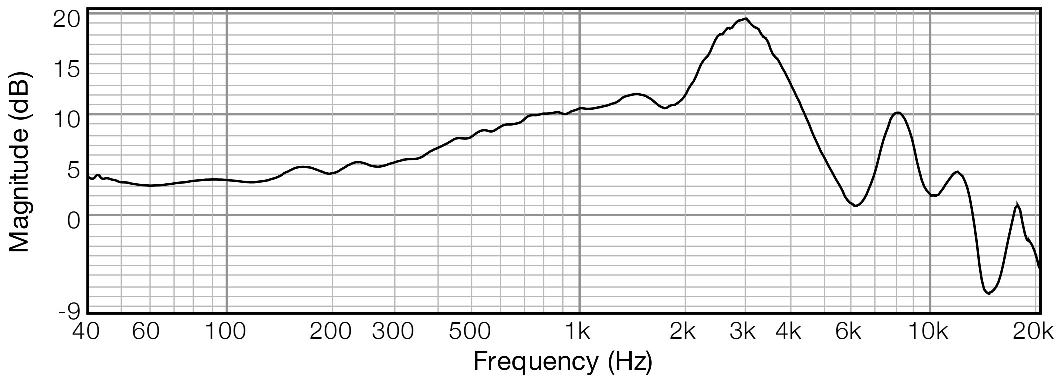

Figure 3.4: Ear canal resonances measured approximately 1 cm deep from the ear canal. The quarter-wave resonance, created by the open ear canal, appears at the frequency of 3 kHz

3.2.2 Binaural Recording

Most of the recordings sold these days are meant for stereo or multi-channel speakers, not for headphones. It is a fact that these kind on recordings are not correct for playback via headphones. When a normal stereo recording (e.g., normal music CD) is reproduced through headphones, the music seems to localize inside listener's head (see Section 2.2.4). This can be avoided effectively if the recording is done with a method called binaural recording.

The binaural recording technique relies on the fact that our brains create the hearing impression based on the two auditory signals received from our right and left ear drum [14]. Binaural recording is usually done with the help of a dummy head replicating human

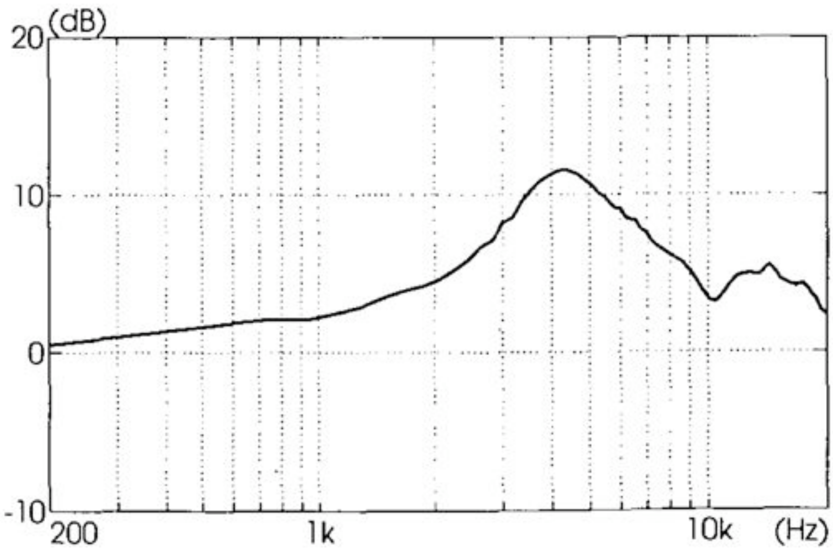

Figure 3.5: The ideal "combined field" frequency response to headphones [25].

features or even with real person wearing a set of binaural microphones in his/hers ears. One example of the dummy head is the Brüel&Kjær's HATS 4128C head and torso simulator. HATS 4128C provides accurate acoustic representation of the average human adult [6]. It has two microphones placed inside its ears and it is also used for measuring headphones (e.g. ARA Headset).

Thanks to binaural recording the localization and the feel of the ambient space is preserved exceptionally well. That is because the binaural recording technique preserves the HRTFs (see Section 2.2.4). The reproduction of the binaural recording does not require any special equipment besides stereo headphones. For example, a dummy head with binaural microphones can be placed on the seat of the director in a concert hall and record a theatrical performance binaurally, thus providing the audio from the performance to everybody with a pair of headphones as if they were sitting on the best place in the concert hall.

The downside is that the generic HRTFs recorded with a dummy head do not fit well for all users. Only those users whose own HRTFs are similar enough compared to the HRTFs of the dummy head can truly feel the benefits of the recordings. The best way to do the recordings would be to use the user's own head with binaural microphones, i.e., real-head recording. This way the individual HRTFs can be preserved.

3.2.3 Leakage of the Headphones

Depending on the type of headphones, different amounts of ambient sounds are transmitted to the ear canal as leakage around and through the headset. In this study we concentrate mostly on the leakages of an in-ear type headphones used in the ARA headset. In some cases headphone leakage can be desirable if one needs to hear the surrounding environmental sounds as well. However, in ARA headset leakages, especially uncontrolled leakages, are harmful. Leakages color the sound signals coming to one's eardrum, since the sound signal that reaches the ear drum is the sum of the pseudoacoustic sound reproduced by the transducer and the sounds leaked from the surrounding environment. Leakages influence mostly at low and middle frequencies. Uncontrolled leakages deteriorate the pseudoacoustic experience [34].

If the leakage paths are known, it is possible to compensate them, e.g., with the help of the equalization of the ARA mixer (see Section 3.3.2). The problem is that different headphones have very different types of leakage and even with the in-ear type headphones, which have the most controllable leakage behaviors, every time the headphone is put into the ear, the fitting is slightly different, therefore, the leakage paths and levels also differ. Leakage is also happening in two directions; from the surrounding environment to the ear canal and from the ear canal to the surrounding environment. The latter case is important because of the pressure chamber principle (see the next Section) [29].

The possible leakage paths are: the leak between skin and the cushion of the headset, the leak through the headphone, bone and tissue conduction, and the vibration of the headphone. Even though in-ear headphones have usually well-fitting silicone ear pads, there can be some leakage between skin and ear pads, especially if the ear pads do not fit well. If the ear pads are the right size for the user, the attenuation of these leakages can be very effective. Most of the in-ear headphone manufacturers provide three different sizes of ear pads, so there is a good chance that the headphones will fit tightly for the majority of users. There are often some kind of small holes in the casing of the headphones, e.g., bass tubes, that allow sound waves to leak through the headphones. The bone and tissue conduction is always present. It is mainly an inconvenience regarding to headphone listening, excluding the special bone conduction headsets. The combination of the bone and tissue conduction and in-ear headphones creates an unnatural sounding effect of speaker's own voice called the occlusion effect (see Section 3.2.4). The headphone itself mechanically transmits sound both inside out and outside in. The vibrations of the headphone can alter the fitting of the headphone and also cause some mechanical sounds into the ear canal.

The Pressure Chamber Principle

When listening with normal loudspeakers the sound pressure is produced to the whole room around the listener. When using headphones, especially in-ear headphones, the volume where the sound needs to be produced is very small compared to the volume of a room. The cavity between the in-ear headphone and the ear drum is somewhere around 1 cm^3. With a cavity that small it is easy to produce high sound pressure levels. With low and middle frequencies, where the wavelength is large compared to the earphone driver, the pressure inside the ear canal is in phase with the volume displacement of the transducer membrane and its amplitude is proportional to it [8]. Therefore, the pressure chamber principle enhances the low and middle frequencies. In principle, the headphone should be very tightly fit so there would be no leaks but in reality small leaks do not interfere that much.

The effect of the pressure chamber principle can be seen in Figure 3.6. The low frequencies attenuate as the volume, where the transducer must create the sound pressure, becomes larger.

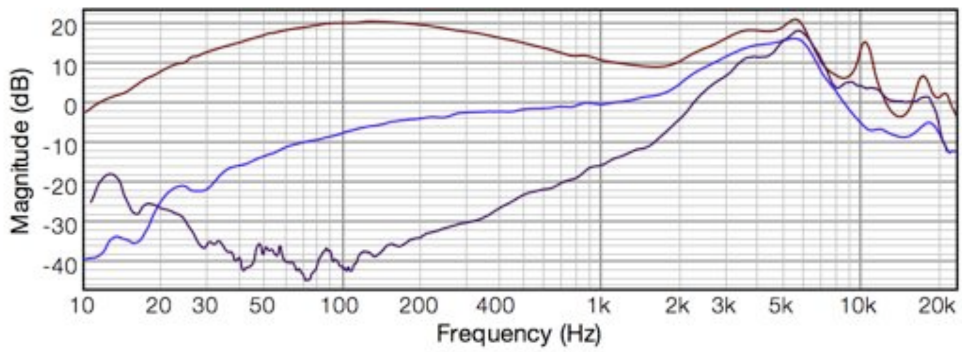

Figure 3.6: Frequency responses of the Philips SHN2500 transducers, used in ARA headset. Highest curve is measured in the ear canal simulator, the middle curve is measured in a resistive-load tube and the lowest curve is measured as free-field radiator (after [30]).

3.2.4 The Occlusion Effect

Auditory occlusion is a phenomenon where in completely (or partially) blocked ears one's own voice sounds hollow and barrel-like. Usually this blockage is some kind of hearing aid device or an in-ear headphone, like in this case, an ARA headset. The occlusion effect is the result of an amplification of one's own voice especially at low frequencies. It occurs due to the bone and tissue conduction. In normal (open ear) situation the bone conducted low-frequency sounds are able to escape from the ear canal but when the ear canal is blocked the low-frequency sounds are trapped inside.

The occlusion effect is subjectively experienced as annoying and unnatural. Hence, it is an unwanted phenomenon and there are two main approaches to reduce the occlusion effect; active and passive reduction. Passive reduction is usually done with vented ear-molds in order to make an escape path for the low frequency sounds. Unfortunately, the downside of this kind of implementation is that the vent also creates leakage path into the ear canal, which creates another kind of problems (see Section 3.2.3). The active implementation uses an in-ear microphone, an earphone driver and an active feedback connection between the earphone driver and internal microphone [5] [23]. With an active feedback system vented ear-molds are not needed, therefore, it would be possible to implement it to the ARA headset as well.

3.3 ARA Hardware

The ARA hardware has been specially designed and built for this purpose [31]. It consists of the ARA headset and the ARA mixer (see Sections 3.3.1 and 3.3.2). The main idea is that the ARA headset would be acoustically transparent (i.e., that everything would sound exactly the same as without the headset), however, this is not the case, thus an ARA mixer is needed for the equalization. Equalization aspires to make the ARA headset as acoustically transparent as possible. With these equipment it is possible to create a good pesudoacoustic representation of the real sound environment. The ARA mixer is also used to add virtual sound objects into the user's sound environment, after which all the necessary equipment for creating an augmented reality audio is in place.

3.3.1 ARA Headset

The main goal for the ARA headset is that it must be able to reproduce the surrounding sound environment. In order to do that the headset has two external microphones, in addition to the earphone drivers. The quality of reproduction must be good enough for allowing the users to wear the ARA headset for long periods of time non-stop. Previous ARA headsets have mainly been loosely-fit in-ear headphones with a manually fitted miniature microphone, which do not entirely fulfill the quality requirements (see, e.g., [16]). The present prototype of ARA headset utilizes the existing Active Noise Canceling (ANC) in-ear headphones. The advantage of the ANC-headphones is that they already have integrated binaural microphones.

The headphones used in the ARA headset is Philips SHN2500 active noise canceling headphones (see Figure 3.7). The headset contains two tightly-fit in-ear headphones (ear

Figure 3.7: Philips SHN2500 headset and ANC control unit [33]

plugs) with integrated microphones. The ANC control unit contains signal processing electronics whose function is to reduce unwanted external noise. The microphone signals are routed to the control unit which recognizes the noise. After that the control unit sends the same noise into the headphones in an invert phase, thus reducing the unwanted noise in the ear canal.

In order to make the headset suitable for ARA usage some modifications have been done. The control unit box is removed and a stereo-mini plug (3.5 mm) is connected to the microphones. So now the headset has two stereo-mini plugs, which can be connected to the ARA mixer, one for the headphones and one for the microphones.

The Properties Of Philips SHN2500

The ARA headset must have good quality components in order to produce the copy of the surrounding environment (pseudoacoustic representation) as unaltered as possible. According to the preliminary measurements [30] the microphones in Philips headset are surprisingly good. In comparison with a high-quality Brüel&Kjær's 4191 free-field microphone the frequency responses are very similar up to about 15 kHz and even after that the responses do not differ dramatically. The directivity of the headset microphones starts to increase somewhere after 6 kHz (see Figure 4.10), while the microphone is turned away from the sound source. That is because the wavelength of the sound signal starts to approach the dimensions of the headphone capsule and the capsule starts to shadow the sound signal more and more.

The earphone driver of the Philips headset also affects the overall quality of the ARA headset. The frequency response of the Philips SHN2500 headphone measured in the resistive-load tube is shown in Figure 3.6 (the middle curve). This frequency response has the same characteristics as the frequency response in Figure 3.5, which depicted an ideal response for headphones. The resistive-load tube offers a seemingly endless transmission channel for the sound waves. That way the tube itself does not create any additional resonances (see more in Section 4.1). The highest peak in Figure 3.6 that is between 5 and 6 kHz is a characteristic feature of the Philips SHN2500 headphone.

3.3.2 ARA Mixer

ARA mixer is the control unit of the ARA hardware [30]. It provides all the connectors and signal processing electronics needed in ARA applications. It contains the preamplifiers for the microphones and amplifiers for the headphones, thus providing adjustable gain. The considerable improvement against the previous ARA mixers is the equalization possibility. Equalization allows the compensation of the acoustical changes created by the ARA headset discussed in Section 3.2. It is also possible to create individual equalizations through measurements in order to compensate structural differences among the variety of ears.

Equalization

The ARA mixer has an adjustable analog equalization circuit [30]. The equalization is implemented by analog electronics because its low latency properties. Digital signal processing causes delays up to milliseconds, which is unacceptable for ARA applications. Low latency is critical because of the headphone leakage. As mentioned in Section 3.2.3, the sound event perceived at the ear drum is the sum of the pseudoacoustic sounds and the leakage that has gone through the headphones. If there is a latency difference between these two, it creates a comb filtering effect, which colors the sound reaching the ear drum by attenuating some frequencies and amplifying the others.

The equalization consists of three adjustable filtering sections for each channel, which are shown in Figure 3.8. First there is a first-order high-pass filter compensating the boost of the low frequencies. The filter is constructed with a simple RC-circuit consisting of a resistor, a potentiometer and a capacitor. The cut-off frequency can be adjusted in the range of 6 to 720 Hz. The latter two filter sections are meant to compensate the resonance problems created by the quarter- and half-wave resonances. The filters are constructed with active LRC-filters, where the inductor is replaced with a gyrator. The first filter is meant to be used as a peak in order to create the missing quarter-wave resonance and the second filter should be used as a notch in order to cancel the half-wave resonance created by the closed ear canal. Both of the filters have adjustable center frequencies, gain, and Q-values. The Q-values have an effect on the bandwidth and the steepness of the peak or notch. If the

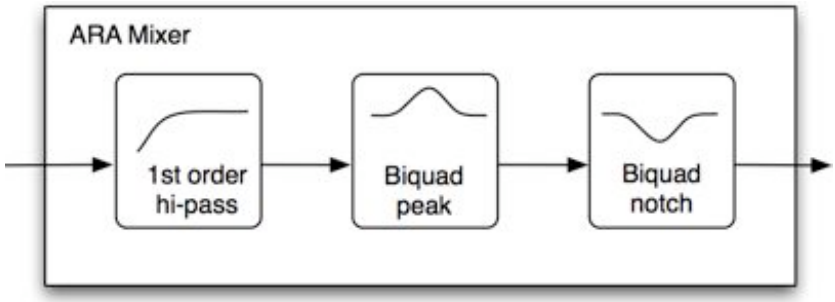

Figure 3.8: Adjustable filter sections in the ARA mixer's equalization circuit [30].

Q-value is high the peak or notch is steep and narrow, and on the other hand, if the Q-value is low the peak or notch is flat and wide. The center frequency of the peak can be adjusted between 700 - 3200 Hz and the center frequency of the notch can be adjusted between 1800 - 8500 Hz [30].

The advantage of the adjustable equalization is that it makes the ARA mixer more generic. It can be used with different types of ARA headsets and it also allows the use of individual equalizations. In order to find a proper equalization curve, some ear canal transfer functions needs to be measured. This has been tentatively done in [31]. The measurements were done in an anechoic chamber (at the Department of Signal Processing and Acoustics). The main idea is to measure the free-field frequency response of the ear canal in the case of an open and closed ear and that way obtain the frequency responses of the natural hearing situation and the pseudoacoustic representation through the ARA headset. These responses are shown in Figure 3.9. The aim is to make the pseudoacoustic representation as similar to the natural hearing situation as possible (i.e., acoustically transparent). As can be seen in Figure 3.9, the pseudoacoustic (gray) curve has boosted low frequencies, a lack of the quarter-wavelength resonance and a new half-wavelength resonance (due to the closed ear canal).

The equalization curve can be obtained by taking the difference between the open ear and the pseudoacoustic case in dB scale. After that the equalization curve must be adjusted to the mixer by hand (with the help of some measurements) [30]. Furthermore, a generic equalization curve was constructed from the average of four individual equalization curves, see Figure 3.10. Figure 3.11 shows the same measurement as in Figure 3.9 except that the ARA headset now uses individually adjusted equalization. The difference between equalized and unequalized pseudoacoustic responses is apparent. The compensation at low frequencies is almost perfect and the resonances are also drawn nearer to the normal hearing situation [30].

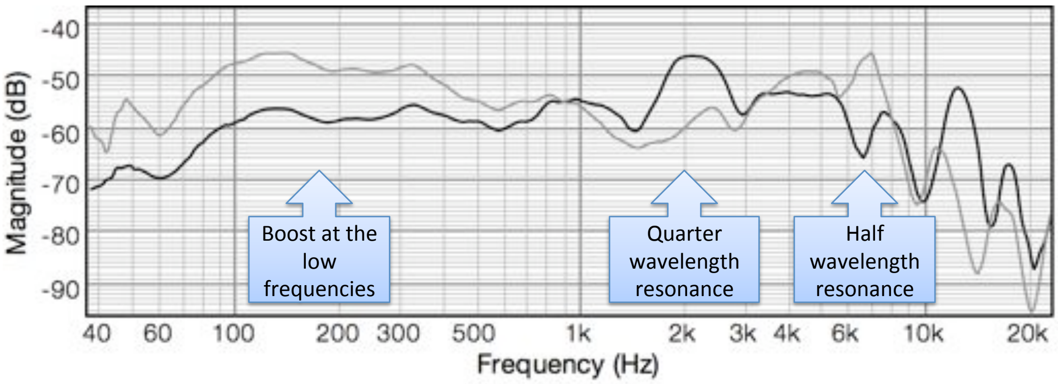

Figure 3.9: Transfer functions from sound source to ear canal: black curve depicts an open ear case and gray curve is measured when the sound travels through unequalized ARA headset into the ear canal (unequalized pseudoacoustic representation) [31].

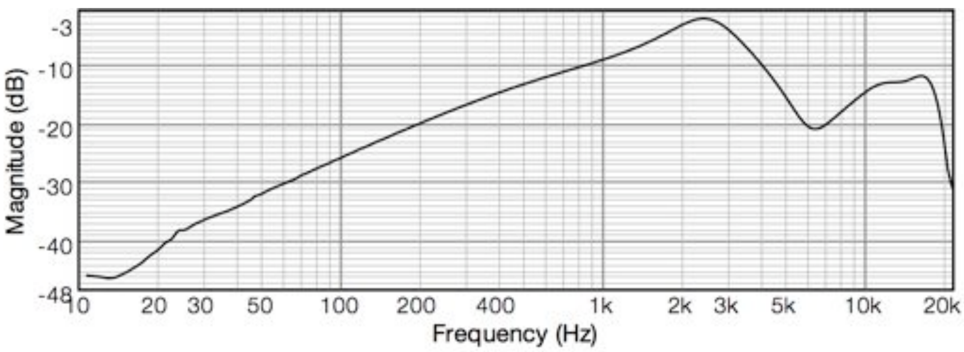

Figure 3.10: Generic equalization curve measured from ARA mixer [30].

Furthermore, a small-scale preliminary listening test was performed [30]. The idea was to test the individual and generic equalizations in comparison to unequalized pseudoacoustic representation. All cases were compared to the natural hearing experience. The result was that both individual and generic equalizations enhanced the pseudoacoustic experience. The number of the test subjects was so small that valid conclusions about which equalization is better could not be made.

Connectors and Controllers

The ARA mixer and its interiors are depicted in Figure 3.12. The ARA mixer uses two 9V batteries as a power supply, which should provide about 20 hours of use. The equalization properties can be adjusted by the 14 adjustable controls located on the circuit board (7

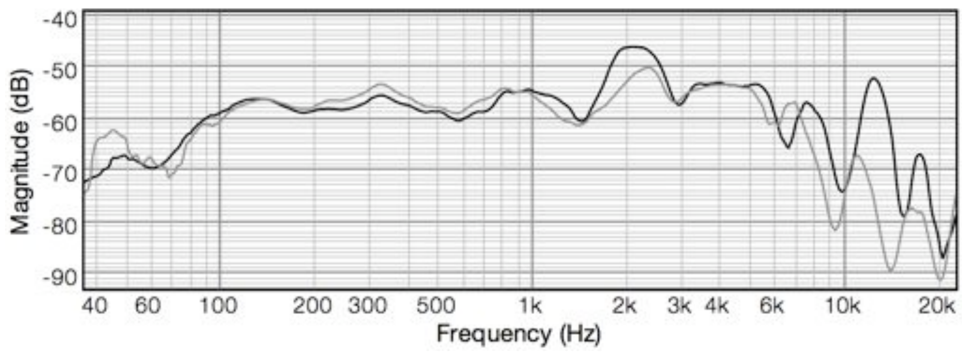

Figure 3.11: Transfer functions from loudspeaker to ear canal: black curve depicts an open ear case and gray curve is measured when the sound travels into the ear canal through the equalized ARA headset (equalized psudoacoustic representation) [31].

controls per channel). The ARA mixer has ten external connectors and controllers; six 3.5 mm stereo-mini plugs, two slide switches and two adjustable controls. The external connectors and controllers are depicted in Appendix A, Figure A.1.

3.3.3 Hands-free for the ARA headset

A hands-free adapter for the ARA headset was constructed in order to enable a cellular phone usage with the ARA headset. The template of the ARA hands-free was a Nokia fashion stereo headset HS-3 [26]. The HS-3 hands-free is connected to a cellular phone with a Nokia Pop-Port™ connector. The most important properties of the Pop-Port™ interface are digital identification, stereo sound, fast data links and voltage supply for the accessories [27]. The original headphones and microphone were removed from the HS-3 circuit board and replaced with two 3.5 mm female stereo-mini plugs. The female mini plug for the headphones was connected directly to the left and right speaker outputs on the circuit board. The left and right channels of the microphone mini plug were connected together with two 2.2 $k\Omega$ resistors in order to convert the stereo microphone signal into a mono signal that can be connected to the circuit board. The ARA hands-free is depicted in Figure 3.13.

New Nokia GSM devices use a 2.5mm/3.5mm interface for audio and video [28]. Accessories such as headphones, video cables and hands-free sets are connected to a mobile phone via a 4-pin plug (see Figure 3.14 and Table 3.1). Figure 3.14 depicts the Nokia AV connector pin layout and the terminal implementation. EMC stands for electromagnetic compatibility and MUX stands for multiplexer.

A basic hands-free headset is composed of a microphone, (stereo) headphones and a

Figure 3.12: The ARA mixer without the cover.

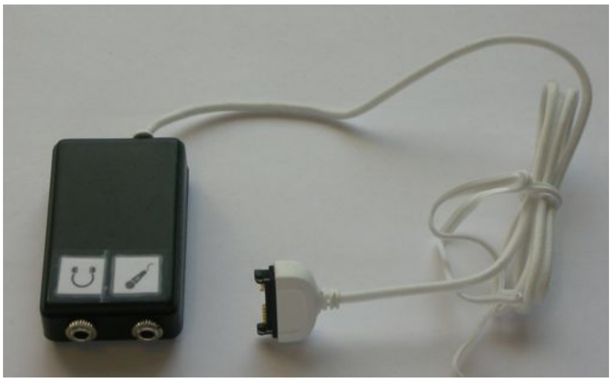

Figure 3.13: The ARA hands-free adapter.

send/end button. The send/end button is implemented with a comparator. While the microphone bias is activated, a press of a send/end button draws the microphone line below the threshold of the comparator and this causes an interruption, which connects or disconnects a call [28].

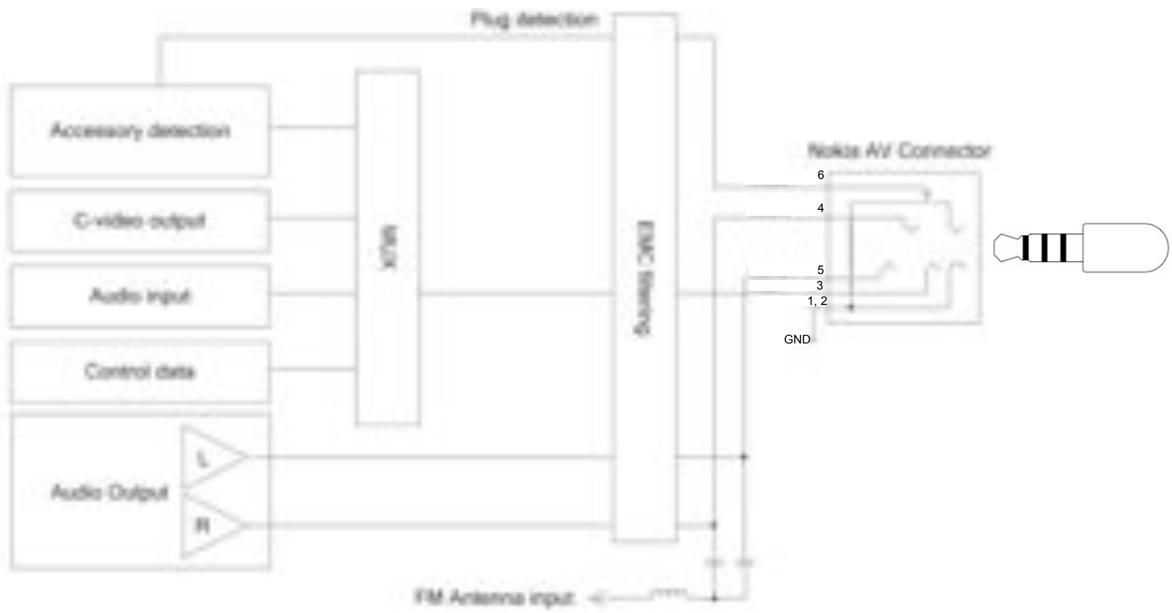

Figure 3.14: Nokia AV connector pin layout and terminal block diagram [28].

Table 3.1: Nokia AV interface signals [28].

Pin number	Description	Direction
1, 2	Ground	-
3	Multiplexed microphone audio and control data, C-video out	In / Out
4	Audio output right channel	Out
5	Audio output left channel	Out
6	Terminal internal connection, plug detection	In

3.4 Head-Tracking and Positioning

Many ARA applications require the information about user's location, orientation and head position. These kind of position-aware ARA applications need the information in order to create practical and realistic virtual sound sources. When using a pair of headphones without applying any head-tracking function, the sound surroundings move as the user's head moves, turns or tilts. This is quite unnatural in most situations and enhances the lateralization effect. This can be avoided with the use of head-tracking. If the orientation of user's head is tracked it is possible to create stationary virtual sound sources, which stay in a fixed position, even if the user turns his/her head (see Figure 3.15). This reduces the lateralization effect because the sound source seems to behave as if it was really out there. The head turning also helps to locate sounds that are coming directly in front or from behind of the subject.

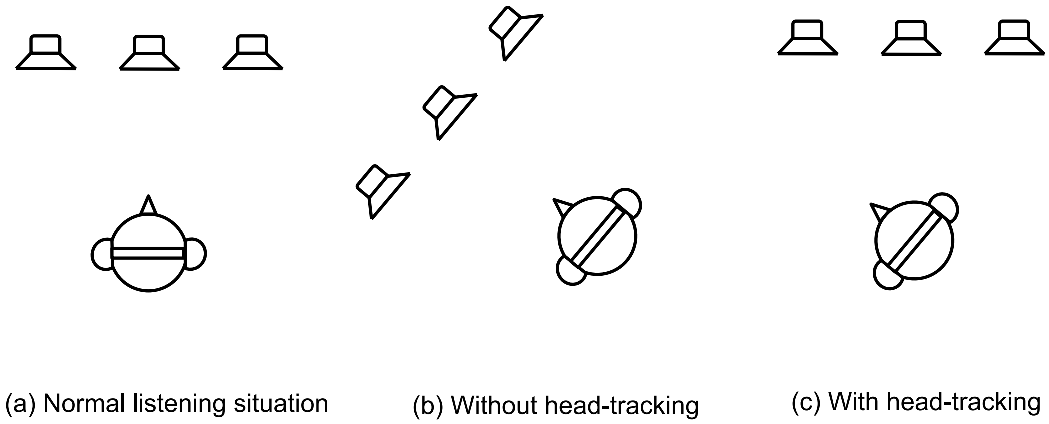

Figure 3.15: (a) is a normal headphone listening situation, where the loudspeakers represent virtual sound sources. (b) and (c) are examples where a virtual sound field moves along with the head without the head-tracking function (b), and where the sound field remains in a fixed position with the help of head-tracking (c)

3.4.1 Positioning

One of the ways to fix the location of the user is with GPS (Global Positioning System). It is nowadays widely used in the navigation systems of cars and even in some mobile phones. A GPS receiver locates itself with the help of a constellation of satellites that orbits the earth. The advantage of GPS is that it can be used worldwide and it is fairly accurate. The downsides are that it is not accurate enough to track the user's head movements and that it requires a line of sight to the satellites in order to operate. Thus, it is only suited for outdoor use and that is not sufficient for ARA applications.

Another common technology that can be used for positioning is WLAN (Wireless Local Area Network). Many current mobile phones and other portable devices already have built-in WLAN. WLAN access points can also be found in many public places, including schools, libraries and airports. Most of WLAN-based positioning is based on signal strength of multiple access points [21]. The advantage of this technique is that it can be used indoors, but similar to GPS it is not accurate enough to track the user's head movement.

3.4.2 Head-Tracking

Head-trackers are used to track the movement of user's head. Positioning a point in 3-D world requires three degrees of freedom (abbreviated 3-DOF), in Cartesian coordinates that means we have to know x, y and z coordinates. If we need to know the orientation of the object as well we need additional three DOFs. These are known as pitch (elevation), yaw (azimuth) and roll (see Figure 3.16). Therefore, we need 6-DOFs in order to track the users location and orientation [39].

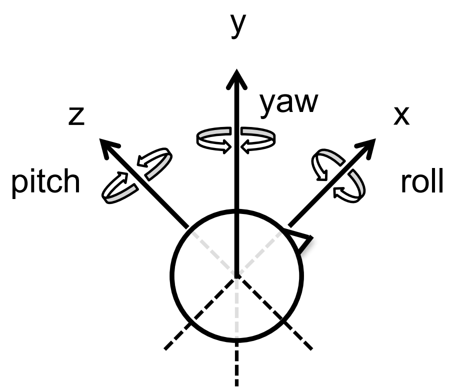

Figure 3.16: The 6-DOFs needed to exactly determine the orientation (pitch, yaw, roll) and the position (x, y, z) of a user.

There are many different kinds of head-tracker implementations. The basic idea in most of them is that there is a transmitter and a receiver. The receiver receives signals from the transmitter and with the help of these signals it determines the position and orientation of the transmitter in relation to itself.

Mechanical Head-Tracking

Mechanical head-tracking is based on a physical connection between the user's head and the receiver. The physical connection can be, e.g., a mechanical arm attached to user's head that transfers the head movements into the receiver. Mechanical head-trackers are accurate and have low latency, but the downsides are the extra gear the user has to wear and their poor mobility.

Electromagnetic Head-Tracking

Electromagnetic head-trackers are based on electromagnetic fields and sensors. The magnetic field is usually created by a set of coils located in a transmitter. A magnetic sensor (a receiver) is attached on the top of user's head and it determines the user's orientation by the strength and angles of the magnetic fields [39]. The positive sides with electromagnetic sensors are that they are fairly accurate, mobile and do not require line of sight between the transmitter and receiver. The downside is the fact that all electrical devices emit electromagnetic radiation, which can cause interference with the desired magnetic field.

Another possibility is to use the magnetic field of the earth as a reference, but the downside with this technique is that only 3-DOF can be tracked (roll, pitch and yaw). This technique coupled with GPS positioning would provide a 6-DOF tracking device, as the GPS takes care of the positioning while the electromagnetic sensors determine the orientation of the user.

Optical and Video-Based Head-Tracking

In optical head-tracking the receiver is some kind of optical device like an ordinary video camera or an infrared camera. Infrared camera is a good choice because then normal (visible) lights do not interfere with it and infrared LEDs (Light Emitting Diode) can be used, which are invisible to the naked eye. The basic idea is that the infrared LEDs send a known signal, which the cameras detect and based on that the position and orientation of the user is computed. The infrared-based head-tracking can be done in many ways. One way is to place the infrared LEDs and cameras in fixed positions somewhere around and above the user and then attach multiple reflective objects (e.g., spheres) on top of the user's head. When the LEDs send an infrared pulse the cameras detect the reflections coming from the reflecting objects.

Another way is to place the cameras somewhere around the user and attach a set of infrared LEDs on the top of user's head. This way the LEDs move as the user's head moves. This requires a power supply in the head-tracker for the LEDs, which the reflecting spheres do not require. The advantages of the optical head-tracking are the high accuracy and low latency (signals travel at the speed of light). The downside is that there must be a line of sight between the user and the cameras.

Video-based tracking is a little different compared to the latter case. It is based on a regular video footage. The main idea is to track something with the help of a video camera. A simple case could be, e.g., to follow a red-hatted man in a crowd, assuming that no one else is wearing a red hat. Video tracking is quite heavy computationally because each frame recorded by the video camera must be analyzed with a chosen algorithm in order to track the desired object. The positive side is that there are many different objects which can be tracked, such as a red hat, face or eyes. There are existing face recognition devices on the market, for example, some still cameras and webcams are able to recognize and track a face with the help of video tracking.

Inertial Head-Tracking

An inertial head-tracker uses the laws of inertia to track the orientation of an object. It does that by the help of accelerometers and gyroscopes [39]. An accelerometer is a device that measures acceleration, whereas a gyroscope can be used for measuring orientation based on the principles of angular momentum. Accelerometers and gyroscopes together provide enough information in order to calculate orientation changes. Inertial-based head-trackers have no range limits or line of sight troubles because there are no stationary receivers that need to be connected. Downsides with inertial trackers are the fact that only 3-DOF can be tracked and that slow movements are hard to track accurately.

Acoustic Head-Tracking

Acoustical head-tracking is based similarly on transmitters and receivers as well as many other head-tracking technologies. Only, in this case, the transmitted signals are acoustical sounds. There are two basic possibilities to construct an acoustical head-tracker. One way is to use the subject as a sound source and the other way is to use stationary anchor sources and use the subject as a receiver. In both cases the tracking is based on the arrival times and level differences of the signals [36]. When using the subject as a sound source there needs to be static microphone arrays installed around the user and the user has to wear some kind of loudspeakers in order to transmit the tracking signals. This technique has some problems concerning the high power consumption of the loudspeakers and the fact that the device attached to user must be as wearable as possible.

The second case where the user wears microphones and there are stationary anchor sources is an ideal case for an ARA environment. This is because the ARA user is already wearing a set of binaural microphones, thus no additional equipment for the user is needed. This is called binaural tracking. Another great advantage in binaural tracking is that the user is wearing the microphones in his/her ears, so it works kind of as an ear-tracker, and it can be applied in ARA applications directly. The stationary anchor sources are usually loudspeakers and the signals transmitted by them can be known or unknown. In fact there are three categories of the anchor sources [36]:

1. Anchor locations, signals and signaling times are known (synchronous tracking),
2. Anchor locations and signals are known (asynchronous tracking),

3. No knowledge of the surrounding sound sources.

If the source signals of the anchors are known, it improves the robustness of the tracking compared to the third case where the signals are unknown. Furthermore, the more anchors there are the more robust the system will be. The unknown signals can be different stationary sound sources around us, such as noise from an air conditioning vent or from the cooling fan of a computer. In this case the tracking and positioning is done by comparing the signals of the binaural microphones.

In an ideal case the anchor signals should not be heard by the users. In other words the anchor signals should be either low- or high-frequency signals that are below the masked threshold of hearing. Background noise creates interferences at low frequencies, thus high frequency anchor signals are preferred [36]. The threshold of hearing also arises towards high frequencies so higher sound pressure levels can be used.

Of course, there are interferences at high frequencies also but the higher sound pressure levels enable better signal-to-noise ratios for the anchor signals. However, the accuracy and latency of acoustic head-trackers are slightly limited because of the limited speed of sound. Furthermore, the speed of sound varies depending on the environmental conditions, e.g., because of temperature changes, that way decreasing the accuracy, and the speed itself is not that fast after all (compare the speed of sound 343 m/s vs. the speed of light 300 000 000 m/s). In addition to this, depending on the type of a room, there are different room reflections that can cause the reception of a reflected signal that could interfere with the head-tracking process.

3.5 Possible Application Scenarios

The ARA technology provides an implementation platform to many new and innovative applications. There are many new applications under development and a lot more is yet to come. Some of these application scenarios are briefly presented in this Section. ARA applications can be categorized in many ways, including communication or information services, human-to-human or human-to-machine communications [22]. For example, binaural telephony and audio meetings (Sections 3.5.1 and 3.5.2) are human-to-human communication services, whereas virtual audio tourist guide (Section 3.5.6) is a human-to-machine information service.

3.5.1 Binaural Telephony

Normal telephones and mobile phones transmit mono sounds and limit their bandwidth to 300 Hz - 3400 Hz. There are some hands-free sets which have two earplugs but they just reproduce a mono signal to each earpiece, thus it is not really a binaural telephony. Binaural telephony means that the binaural signals from both users are transmitted to each other. This cannot be done with normal telephone lines nor GSM networks, thus we need another solution, such as VoIP (Voice over IP). VoIP uses IP packets to carry the binaural telephony signals over the network (Internet). With VoIP there is no need to limit the frequency bandwidth so it is possible to transfer the whole sound surroundings around one user to another user, i.e., the far-end listener hears the same pseudoacoustic reproduction as the near-end user.

The inconvenience in the binaural telephony with the ARA headset is when either of the users talks, the voice is located inside the other user's head and in some situations the voice can be too loud as well. This can be avoided if the user's own voice is detected and then panned around the far-end user with the help of HTRFs [22]. This creates more natural feeling to the conversation because it seems that the far-end user is in front of or around the listener.

3.5.2 Audio Meetings

Audio meetings are very similar to binaural telephony except that there are usually greater amount of participants present at the audio meeting. Nowadays audio meetings are very common because of the globalized businesses, which often makes the face-to-face meetings practically impossible. Traditionally audio meetings are held with the help of telephones and speakerphones. One of the problems here is the lack of telepresence because all the participants are reproduced through one phone or speakerphone. With the ARA headset virtual meetings are brought to a new extent. It is easy to form discussion groups that consist of remote and local participants. Remote participants can be panned around the user (see Figure 3.17) and blended to the same acoustical environment as the user. This way it is much easier to separate the participants and distinguish who is talking [18].

There are many ways to utilize the ARA technology in different type of audio meetings. One scenario could be that a traditional meeting is arranged and some team member is out of town. If he/she has an ARA headset and at least one person who is present at the meeting has an ARA headset, the out-of-town team member can participate the meeting virtually (see Figure 3.18). Because of the ARA technology, the out-of-town employee can hear exactly the same what the person who wears the ARA headset in the meeting room hears. One downside is that if there is not any speaker system in the meeting room only the person who wears the ARA headset can hear the other team member.

Another inconvenience, similar to the binaural telephony, is that the near-end user's own voice localizes inside the far end user's head and that the voice is much louder than the voices of the other participants at the meeting room. This can be avoided using the same

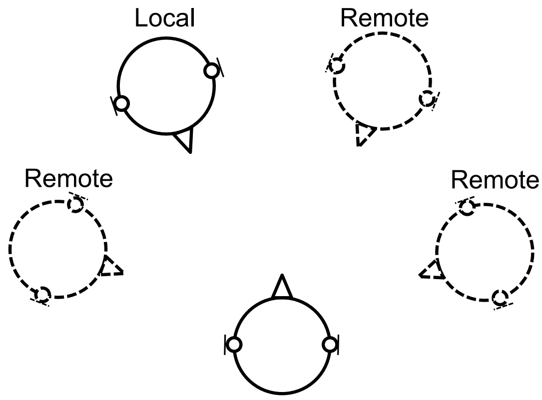

Figure 3.17: Diagram of audio meeting using ARA technology where participants are panned around the user.

principle as with the binaural telephony.

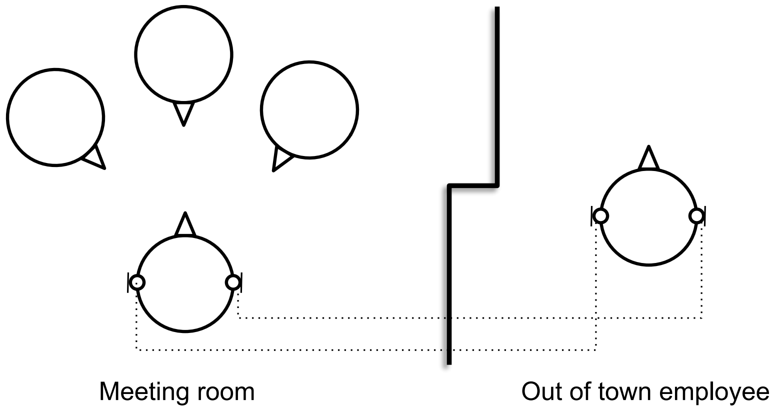

Figure 3.18: Diagram of audio meeting where an out-of-town employee is participating the meeting with the help of ARA headset.

3.5.3 Acoustic Post-It Stickers

Acoustic Post-it sticker application is a location-based communication application [22]. The basic idea of this application is that users can leave auditory messages to a specific location or attach them to objects. The messages can be left to oneself or to other users. For example, if a person is going to a meeting he/she can leave an acoustic Post-it sticker attached to the office door, which says how long the meeting is going to take and where it is held. Now when another ARA headset user comes by the door the message is played back through his/her ARA headset.

3.5.4 Calendar

In audio calendar application the idea is to pan an audio calendar horizontally around the user [38] [22]. The user can record calendar notes, which are then placed to some direction in space according to its time stamp. For example, 9 a.m. could be on the left side of the user, noon in front of the user, 3 p.m. on the right side of the user and 6 p.m. behind the user [38] (see Figure 3.19).

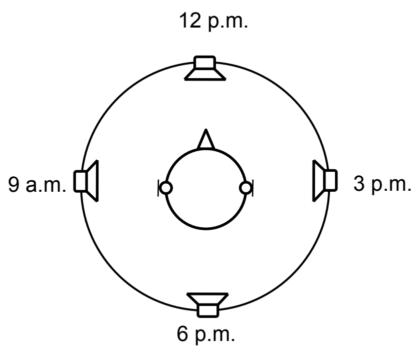

Figure 3.19: Audio calendar application

3.5.5 Audio Memo and Binaural Recording

The ARA headset can be used as a recording device as well. The user can record notes like with an ordinary dictating machine. Recording is done binaurally due to the ARA headset. With binaural recording the surrounding sound environments can be recorded very nicely (see Section 3.2.2).

The idea behind the audio memo application is that all the audio going through the ARA headset is recorded (at least for temporarily). Furthermore, the user can put markers to the recording at any time. This is very convenient, for example, in a situation where someone is introducing himself. The ARA user can put a marker to the point where the other person introduced himself so he/she will not forget his name.

3.5.6 Virtual Audio Tourist Guide

Virtual audio tourist guide is a very interesting idea for an ARA application. It could practically replace the tourist guides and give the user the freedom to explore a city by himself without predetermined routes or schedules. The idea is that the ARA headset has a positioning capability (e.g., GPS, see section 3.4.1), and perhaps a head-tracker as well, thus the application knows where the user is. Then the user can walk around the city with the ARA headset and automatically hear information about interesting places he/she visits. Furthermore, the virtual audio tourist guide could have a GPS navigator application to guide the user from one place to another as well as contain all types of information about the city, such as restaurants, public transportation and concerts [18].

3.5.7 Audio-Guided Shopping

Audio-guided shopping is quite similar to virtual audio tourist guide. The idea is that the user gets product information and pricing about the products he/she is interested in. The user could, for example, stand in front of some shop window and the GPS positioning knows where the user is standing and gives information about that particular shop. Furthermore, advertisements about nearby shops could be provided if the user allows it [18].

3.5.8 ARA Headset Usage in a Car

ARA headset can be used as an auditory navigator similar as the voice guidance in the existing navigators. This of course requires the GPS functionality in the ARA headset. It could also provide additional information, such as traffic announcements, weather forecasts and information about nearby gas stations. What is great about the ARA headset usage in a car is the fact that the ARA headset is hear-through. This allows the user to hear all the important natural sounds related to driving like the engine sounds and traffic, he/she can also communicate with other passengers and even listen to the music that is playing in the background.

A car is also a good environment to implement acoustic head-tracking because there is usually a loudspeaker system installed in the car, which can be used as static anchor sources for head-tracking (see Section 3.4.2). Head-tracking can be used to improve safe driving. For example, if the driver is looking too long somewhere else than the direction of the road, the ARA headset could inform the driver to draw attention to the driving. Furthermore, the system could wake up the driver if it notices that the driver is asleep. Hands-free is also an important part of safe driving and it can be implemented to the ARA headset in a natural way (see Section 3.3.3).

Chapter 4

Laboratory Measurements

4.1 Measurement Techniques for Headphones

There are a number of different ways to measure headphones. The measurement technique used depends on what the objectives of the measurements are. In many cases some kind of an artificial ear is used. When measuring circumaural headphones a dummy head is often a very useful instrument because the headphones are placed around the ear against the head. Usually both artificial ears and dummy heads have artificial ear canals with microphones at the end of the ear canal, where the ear drum is supposed to be. This thesis is concentrated on in-ear headphones as mentioned earlier. When in-ear headphones are measured, often some kind of tube, where the headphone fits tightly, is required (excluding free-field measurements). Some of these measurement devices are presented in the next section.

One common way to measure an impulse response is to utilize a swept-sine technique [13]. The swept-sine technique is also suitable for distortion measurements. The basic idea of impulse response measurements is to apply a known excitation signal $x(t)$ to a system that is being measured, and to measure the response of the system $y(t)$. The excitation signal $x(t)$ is usually wide-band noise or, in this case, sine sweep. Sine sweep is a signal that has a sinusoidal shape with exponentially increasing frequency. The impulse response $h(t)$ can be derived from $x(t)$ and $y(t)$ with deconvolution

$$h(t) = \mathcal{F}^{-1}\left\{\frac{\mathcal{F}(y(t))}{\mathcal{F}(x(t))}\right\}, \tag{4.1}$$

where $\mathcal{F}$ and $\mathcal{F}^{-1}$ depict Fourier transform and inverse Fourier transform, respectively.

4.1.1 Simple Ear Canal Simulator

The ear canal simulator used in this case is a 25 mm long rigid tube with a diameter of 9 mm. It is about the same size as a typical human ear canal. The headphone fits tightly in the other end of the tube and the other end is blocked with a hard wall and tightly inserted Sennheiser KE-4 electret microphone. With this measurement device the ear canal resonances (see Section 3.2.1) are included in the measurement results. In conclusion, the ear canal simulator simulates a situation where an in-ear headphone is inserted into a real human ear. The setup is depicted in Figure 4.1.

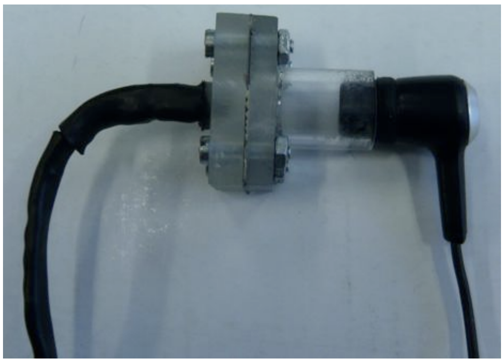

Figure 4.1: The headphone (on the right) and the Sennheiser KE-4 microphone (on the left) are attached to the simple ear canal simulator (in the middle) [30].

4.1.2 Adjustable Ear Canal Simulator

Adjustable Ear Canal Simulator (abbreviated ADECS) is similar to the simple ear canal simulator but in addition it has an adjustable length of the ear canal and microphone position [17]. The range for the length of the ear canal is 0 - 39 mm and the range for the microphone position is 0 - 57 mm starting from the piston (ear drum). The diagram of ADECS is shown in Figure 4.2.

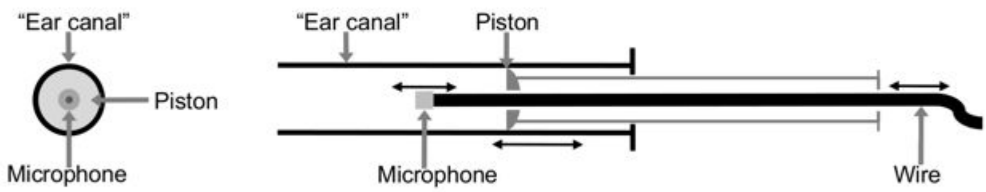

Figure 4.2: The adjustable ear canal simulator [17].

4.1.3 Resistive-Load Tube

Resistive-load tube is about 12 meters long plastic tube with a diameter of 9 mm. The headphone is inserted into one end of the tube and the other end is closed off with cotton wool in order to eliminate reflections from the end of the tube. The main point of the resistive-load tube is that it does not create any additional resonances or reflections. This way the measurement result reflects the measured device, not the characteristics of the tube. The tube offers a seemingly endless transmission channel for the sound waves. It has an approximately resistive impedance equal to the wave impedance

$$Z_w = \frac{\rho c}{A}, \tag{4.2}$$

where ρ is the density of the air (1.15 kg/m^3 at room temperature), c is the speed of sound, and A is the cross-sectional area of the tube [32].

A Knowles FG-3329 miniature microphone is seamed tightly in the wall of the tube, e.g., 3 cm from the end where the headphone is inserted (approximately where the ear drum should be). The measurement results are quite similar to the ones obtained with the ear canal simulators excluding the ear canal resonances and low frequency amplification due to the pressure chamber effect (see section 3.2.3). The resistive-load tube with a headphone inserted into the other end is depicted in Figure 4.3.

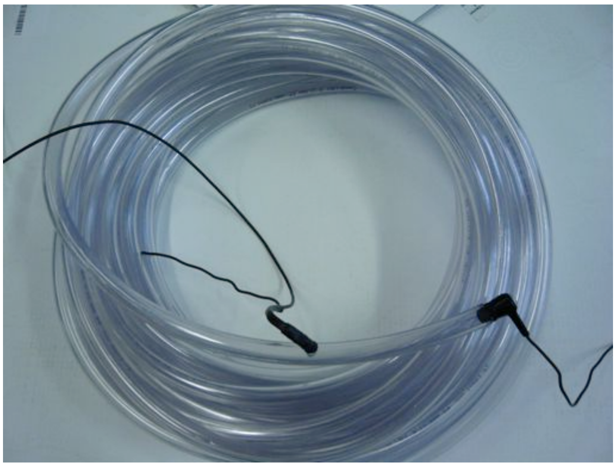

Figure 4.3: The resistive-load tube

4.1.4 Free-Field Measurements

Free-field measurements are done in an anechoic chamber. The headphones can be measured directly or the headphone can be fitted through a hole in a large baffle so it radiates into a half-space (like in loudspeaker element measurements). In free-field conditions low frequencies attenuate almost completely, because there is no cavity where the pressure chamber effect could occur (see Section 3.2.3). Otherwise the measurement results in free-field are fairly similar compared to the matched-impedance tube case.

4.1.5 Philips Earphone with Mounted In-Ear Microphone

A Knowles FG-3329 miniature microphone was permanently installed in front of the Philips SHN2500 headphone [17]. This microphone functions as an in-ear microphone. The idea is that the microphone is always kept in a fixed position, because the microphone position affects the measurement results (see Section 4.6.2). In a previous prototype the in-ear microphone was removable, hence the position of the microphone varied every time it was reinstalled [30]. The in-ear microphone also suits in active occlusion cancellation (see Section 3.2.4). A pair of Philips earphones with mounted in-ear microphones is depicted in Figure 4.4.

Figure 4.4: A pair of Philips earphones with mounted in-ear microphones [17]. The headphones are similar except that the rubber ear cap of the right-hand side headphone has been removed in order to bring out the in-ear microphone.

4.1.6 Headphone Adapter for Brüel&Kjær's Microphone

An adapter for the Brüel&Kjær's 4191 precision microphone was constructed. The adapter allows the usage of a high-quality Brüel&Kjær's microphone in headphone measurements. The other end of the adapter fits very tightly around the Brüel&Kjær's microphone and the other end has a hole of 10 mm in diameter. Different lengths of tubes can be attached into the hole depending on what is measured. The tubes can simulate different lengths of ear canals or a small cavity can be used in order to avoid the resonances of the tube interfering with the measurements. The adapter with a 27 mm tube and a Philips earphone is depicted in Figure 4.5.

Figure 4.5: Headphone adapter attached for Brüel&Kjær's 4191 Microphone.

4.2 Measurements of the ARA Hardware

This section covers the detailed laboratory measurements of the ARA hardware. The headset measurements are composed of earphone driver measurements and headset microphone measurements. One additional measurement was conducted concerning the ARA mixer. Furthermore, some supplementary headphone measurements were also conducted.

4.3 Earphone Drivers

4.3.1 Nonlinear Distortion of the Earphone Drivers

Distortion expresses the impurity of a reproduced signal. Nonlinear distortion is usually unwanted, excluding, e.g., electric guitar where the distortion is expedient. There are different measures that are used to describe distortion. The most commonly used method is to measure the harmonic distortion. Total harmonic distortion (abbreviated THD) is used to measure the percentage of the harmonic distortion. THD percentage can be calculated with

$$THD = \frac{\sqrt{{U_1}^2 + {U_2}^2 + {U_3}^2 + ...}}{\sqrt{{U_0}^2 + {U_1}^2 + {U_2}^2 + {U_3}^2 + ...}} \cdot 100\%, \tag{4.3}$$

where U_0 is the amplitude of the fundamental component, U_1 is the amplitude of the first harmonic component, U_2 is the amplitude of the second harmonic component and so on. The harmonic distortion emerges due to the nonlinearity of headphone elements. The harmonic frequencies are integral multiples of the input frequency, $2f, 3f, 4f$ and so forth.

Small amount of distortion ($< 1\% = -40$ dB at low harmonics) does not deteriorate subjective sound quality [8].

To find the distortion levels of the Philips SHN2500 earphones a set of measurements were conducted. The goal was to specify the harmonic distortion of the Philips headset. Six pairs of Philips SHN2500 headphones were measured in order to find out the variation between different headphone units.

Measurement Setup

The setup for distortion measurements included a Brüel&Kjær's 4191 precision microphone, a Nexus type 2690 pre-amplifier, the headphone adapter for Brüel&Kjær's microphone (introduced in Section 4.1.6) and an Edirol FA-101 firewire audio interface. The Brüel&Kjær's 4191 microphone was connected to the Nexus pre-amplifier, which was adjusted to sensitivity of 10 mV/Pa. The output of the Nexus was connected to the Edirol FA-101 audio interface. The headphone adapter and a tube with a length of 27 mm and diameter of 9 mm was fitted around the Brüel&Kjær's microphone (see Figure 4.5). The cavity of the tube simulates the cavity of an actual ear canal.

Matlab with Playrec (multi-channel Matlab audio) utility was used as a measurement software [3][4]. A logarithmic amplitude ramp was generated with Matlab and played through the Philips SHN2500 headphone, which was attached to the Brüel&Kjær's 4191 microphone via the headphone adapter. The Matlab script calculates the magnitudes of the second and third harmonic components at input frequencies of 20 Hz, 50 Hz, 80 Hz, 150 Hz and 500 Hz. The used sound pressure levels were between 85 - 130 dB.

Measurement Results

The measurements showed that there are no disruptive distortion. Audible distortion occurs only at lowest frequencies with very high sound pressure levels ($>$ 110 dB). The results of the distortion measurements are depicted in Figure 4.6.

4.3.2 Frequency Responses of the Earphone Drivers

Frequency response measurements are a basic way to study the performance of headphones. Frequency response measures the ability of headphones to reproduce different frequencies. Unlike loudspeakers the ideal frequency responses of headphones are not flat (see Figure 3.5). Six pairs of Philips SHN2500 headphones were measured, i.e., twelve frequency responses in total, to see their general behavior and variation between units.

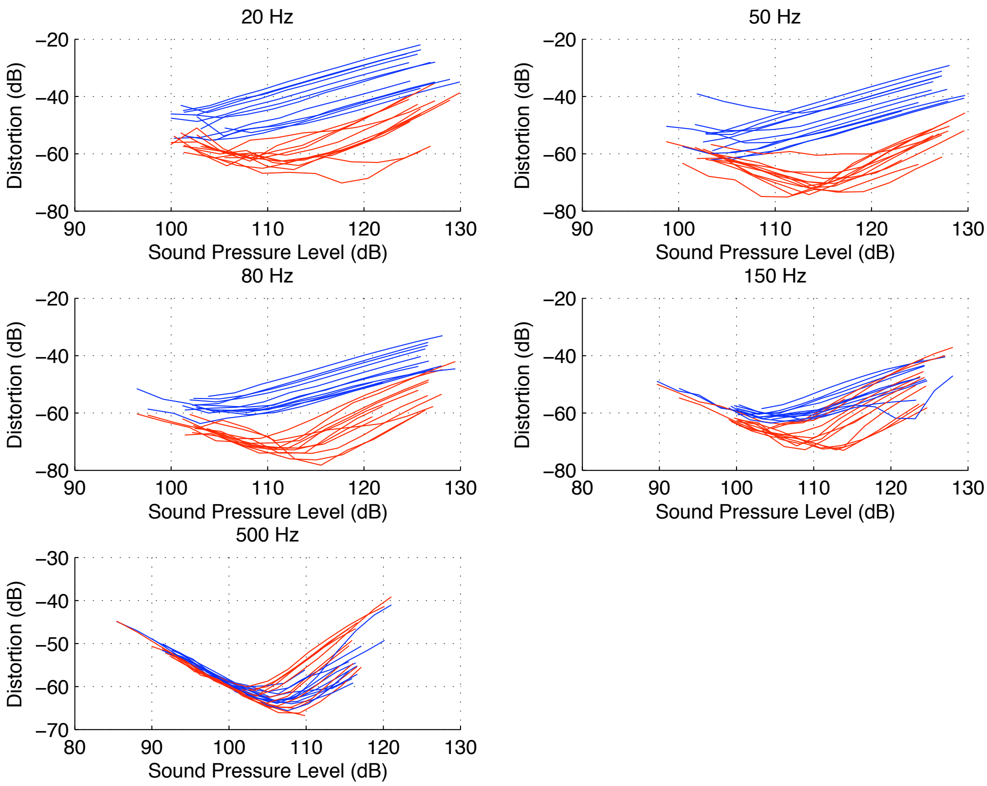

Figure 4.6: Distortion levels of the Philips SHN2500 earphone drivers. Blue curves are 2^{nd} harmonic components and red curves are 3^{rd} harmonic components of different earphone units. At lowest levels the curves turn up, which is not nonlinear distortion but noise in the measurement.

Measurement Setup

The equipment used in the frequency response measurements consisted of a similar setup as used in distortion measurements in Section 4.3.1. The only differences were that the tube used in the headphone adapter was shorter (a length of 9 mm and a diameter of 9 mm) and the input levels were slightly lower. A shorter tube was used in order to avoid the tube resonances influencing the measurement results. The gap between the inserted headphone and Brüel&Kjær's microphone was about 4 mm long and the volume of the cavity was approximately 0.25 cm^3.

Measurement Results

The measurement results can be seen in Figure 4.7. The frequency responses of individual units are quite similar, in general terms, except for a couple of exceptions. The lowest curve (dashed red line) seems to suffer from some leakage. The resonance peaks of the Philips SHN2500 are located approximately at 4 kHz and 6 kHz.

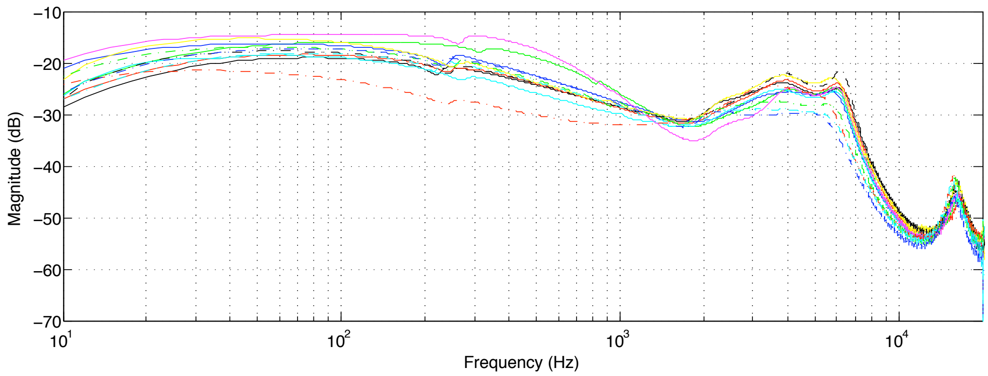

Figure 4.7: The frequency responses of Philips SHN2500 headphones.

4.3.3 Other Earphone Drivers

Four other in-ear headphones from different manufacturers were also measured in order to gain some perspective and comparison with the Philips SHN2500 headphones. These headphones were Sennheiser CX300, Koss KEB24, Nokia in-ear headphones and Sony MDR-EX52 LP.

The frequency responses of these headphones were measured similarly as the frequency response of Philips SHN2500. The results are depicted in Appendix B, Figure B.1. The results show that the frequency responses are quite different in some respects. There are 15 dB differences in bass responses and the resonance peaks are of different shapes and sizes and their location varies from 3 kHz to 7.5 kHz. The optimal Q-value and magnitude of a resonance peak is very individual, because it depends on the characteristics of one's ear canal.

The distortions of the four headphones were also measured the same way as the distortion of the Philips SHN2500. These results are depicted in Appendix B, Figure B.2. As the results imply, the distortion levels of the Philips SHN2500 are relatively low. The Sennheiser CX300 was the only one that had lower distortion levels than the Philips SHN2500.

4.4 Headset Microphones

4.4.1 Frequency Responses of SHN2500 Headset Microphones

Frequency response measures the way a microphone responds to different frequencies. The ideal frequency response for a microphone depends on its purpose of use. A microphone with a flat response responds equally sensitive to all frequencies, hence resulting an accurate reproduction of the original sound environment. A flat and omnidirectional frequency response serves the ARA headset purposes very well, since the main idea in ARA is to reproduce the surrounding sound field as accurately as possible. Eight Philips SHN2500 headphone pairs (16 microphones) were measured.

Measurement Setup

The measurements were performed in the large anechoic chamber at the Department of Signal Processing and Acoustics in Helsinki University of Technology. The measurement setup for the frequency response measurements was the following. A Genelec 8030A loudspeaker was placed 1.6 meters away, directly in front of a microphone. The microphone was connected to an ARA mixer and the signal from the microphone output of the ARA mixer was connected to an Edirol FA-101 firewire audio interface. A Brüel&Kjær's 4191 microphone was used as a reference. The microphones were measured one by one, using Matlab and Playrec software.

Measurement Results

The frequency responses of the microphones are presented in Figure 4.8. As can be seen, all of the frequency responses are very similar. The shapes are very convergent and there is only a 3 dB difference between the most and least sensitive microphone.

The frequency responses in Figure 4.8 contain also the properties of the Genelec 8030A loudspeaker. When deconvolution [24] is performed between the frequency responses in

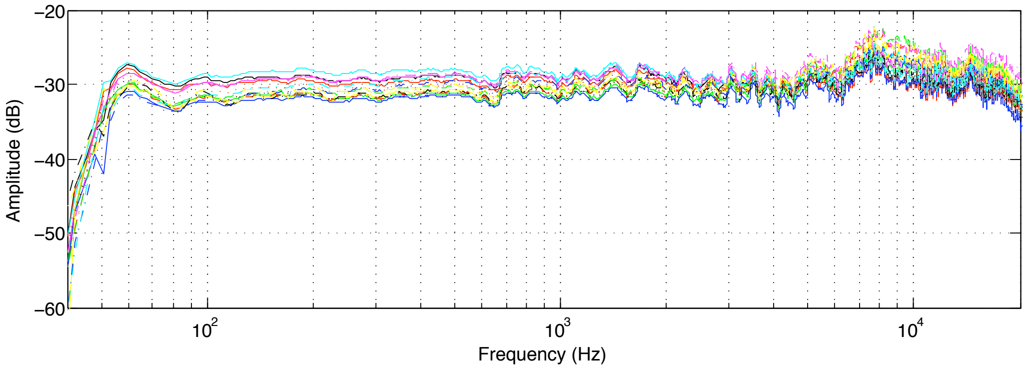

Figure 4.8: Frequency responses of the Philips SHN2500 microphones, including the response of the Genelec 8030A.

Figure 4.8 and the Brüel&Kjær's 4191 reference signal, the effect of the Genelec loudspeaker can be removed. Figure 4.9 shows the deconvolved average frequency response of the microphones. The frequency response in Figure 4.9 is quite flat, which is a good feature concerning the ARA usage. Frequencies above 3 kHz are somewhat boosted (5 dB at the maximum), but this is not a problem because the directivity of the microphone starts to attenuate frequencies after 3.5 kHz (see Section 4.4.3).

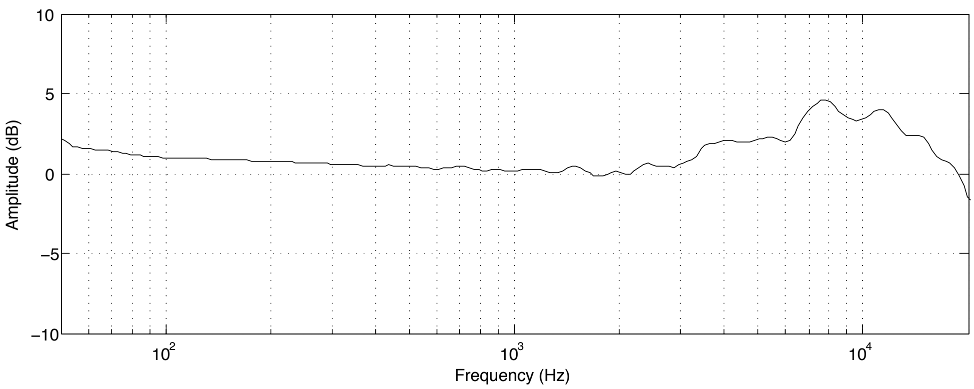

Figure 4.9: Average frequency response of the Philips SHN2500 microphones deconvolved with the Brüel&Kjær's 4191 reference microphone (1/6 Octave Smoothing).

4.4.2 Self-Noise of Headset Microphones

The self-noise of a microphone is the amount of noise that the microphone creates when recording in complete quietness (usually stated as A-weighted value in dB). A-weighting attenuates low and high frequencies and produces a result that is approximately equivalent to what we hear. Furthermore, the self-noise represents the bottom level of the microphone's dynamic range, since it masks sounds that are quieter than that [12].

To find the self-noise levels a set of measurements were conducted in the large anechoic chamber at the Department of Signal Processing and Acoustics in Helsinki University of Technology. The anechoic chamber is a suitable place to measure the self-noise levels because of its quietness (below 0 dB at mid frequencies). The chamber is isolated from the rest of the building in order to insulate the structure-borne noises as well. Four Philips SHN2500 microphones were measured.

Measurement Setup

The measurement setup was basically the same as used in the headset microphone frequency response measurements in Section 4.4.1. Firstly a calibration signal for Brüel&Kjær 4191 was recorded (94 dB, 1000 Hz). Then a reference signal was played with the Genelec loudspeaker (about 49 dB, 1000 Hz), which was recorded with Brüel&Kjær's 4191 microphone and with four Philips SHN2500 microphones. After that some silence was recorded with occluded and open microphones. Audacity software was used to record all signals [1].

Measurement Results

The measurement data was imported to Matlab, where the self-noise levels were calculated. First the RMS value of the calibration signal was calculated. It is known that the RMS value of the calibration signal equals 94 dB. With the help of that information, we can calculate the decibel values of the other recordings from their RMS values. The reference signal of the Brüel&Kjær's microphone was A-weighted, with an A-weighting filter in Matlab, and the decibel value was calculated based on its A-weighted RMS value (48.7 dB). This was compared with the reference signals of the Philips microphones and the levels of the Philips microphones were corrected to meet the same decibel value as the recording made with Brüel&Kjær's 4191 microphone. The average correction was 0.5 dB. After the corrections the self-noise levels were calculated from the RMS values of the Philips SHN2500 recordings. The results are presented in Table 4.1.

Table 4.1: The A-weghted self-noise levels of Philips SHN2500 microphones.

	Open	Occluded
Microphone 1	26.8 dB	26.6 dB
Microphone 2	26.5 dB	25.9 dB
Microphone 3	26.4 dB	25.9 dB
Microphone 4	27.3 dB	26.4 dB
Average	26.7 dB	26.2 dB

4.4.3 Directivity of Headset Microphones

The components of the microphone and the shape of the microphone casing affect the directivity of the microphone. The desired directivity pattern for a microphone depends on the end use of the microphone. In many occasions a well controlled directivity is needed. A simple directivity measurement was performed. The frequency response of the Philips SHN2500 microphone was measured in three different angles (0°, 45°, 90°).

The measurement was conducted in the large anechoic chamber at the Department of Signal Processing and Acoustics in Helsinki University of Technology. The microphone was placed 1.6 meters away from a Genelec 8030A loudspeaker, firstly facing directly toward the Genelec loudspeaker, then in 45° angle and finally in 90° angle. The results are depicted in Figure 4.10.

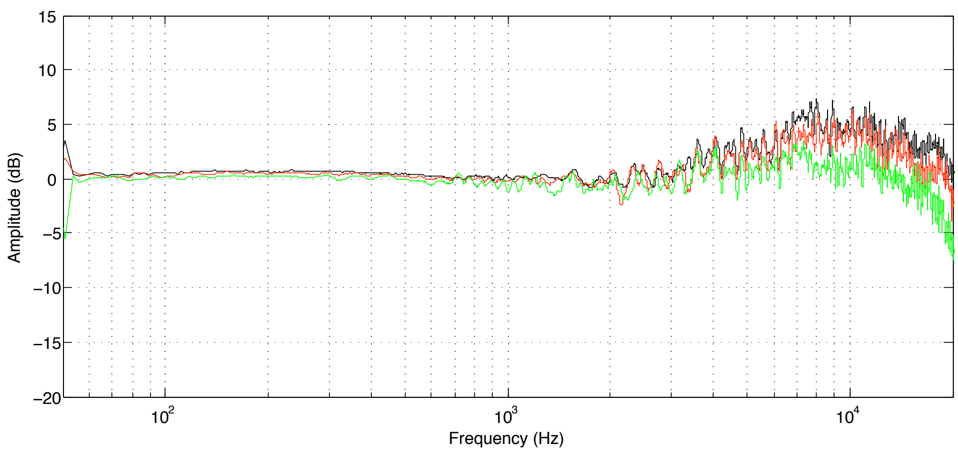

Figure 4.10: The directivity of the Philips SNH2500 microphone in three different angles; 0° (black), 45° (red) and 90° (green). The response of the Genelec loudspeaker has been removed.

4.4.4 Sensitivity of Headset Microphones

The sensitivity of a microphone expresses how the microphone converts air pressure into output voltage. The higher the sensitivity of the microphone is, the less it requires amplification, which may lead to better signal-to-noise ratio. Three different Philips SHN2500 microphones were connected to a 9 V biasing network. The network blocked the DC voltage from the captured signals with a 2 kΩ resistor and a 17 μF capacitor. The output of the biasing network was connected to a Fluke 45 multimeter.

A Genelec 8030A loudspeaker was placed 1.3 meters away from the microphone in an anechoic chamber. 1 kHz signal was played through the Genelec 8030A and recorded with a Brüel&Kjær's 4191 microphone as a refence. The sound pressure level at that point according to the reference signal was 75.5 dB. The same 1 kHz signal was played while the Philips microphones were connected and the RMS voltage values were read from the display of the Fluke 45. Table 4.2 shows the RMS voltage levels obtained from the different Philips microphones and their calculated sensitivities in millivolts per Pascal.

Table 4.2: The RMS voltage levels obtained from the Philips microphones with A-weighted sound pressure level of 75.5 dB at 1 kHz and their sensitivities in millivolts per Pascal.

	RMS voltage	Sensitivity
Microphone 1	5.88 mV	35 mV/Pa
Microphone 2	5.64 mV	34 mV/Pa
Microphone 3	5.75 mV	34 mV/Pa
Average	5.76 mV	34 mV/Pa

4.5 ARA Mixer Measurement

The basic measurements of the ARA mixer were conducted in a previous study [30]. This Section presents an additional channel separation measurement.

4.5.1 Channel Separation

In the case of stereo audio, channel separation represents the leakage of sound from one channel to another. The channel separation is measured by taking the decibel difference between the excitation signal and the leakage from the other channel. This kind of channel leakage can deteriorate the quality of the pseudoacoustic representation, especially the directional information, because the leakage disturbs the natural ILD and ITD cues.

Measurement Setup

In order to find the channel separation of the ARA mixer, a set of measurements were conducted. The channel separation was measured from two different outputs of the ARA mixer with different equalization settings. Both measurements were done so that a tube with a small cavity was attached around the microphone of the left ARA headset and the right channel microphone was occluded. A Sennheiser CX300 headphone was inserted into the other end of the tube and used as a source for the excitation signal. The excitation signal was a two second sine sweep.

First the channel separation was measured from the headphone output. This represents the channel separation actually heard when the ARA headset is used. Second, the channel separation was measured from the microphone output, which is used, for example, for connecting an outside recording device or when transmitting the near-end pseudoacoustic representation to a far-end user. Both the excitation signal from the left channel and the signal leaked to the right channel were measured at the same point, either from the headphone output or the microphone output. Figure 4.11 shows the schematic representation of a single channel of the ARA mixer.

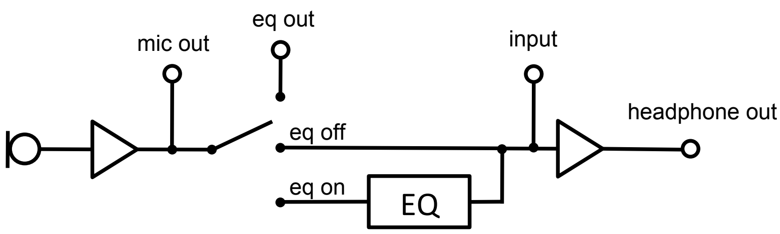

Figure 4.11: Schematic representation of a single channel of the ARA mixer.

The different equalization settings were: equalization on, equalization off, and equalization out (see Figure 4.11). The equalization on setting is the general equalization used in the ARA mixer. The equalization off represents the unequalized pseudoacoustic case. A cable was connected between the equalization out and the input, thus the equalization out setting was similar case as the equalization off. The background noise was measured using the equalization out setting without connecting the equalization out to the input.

Measurement Results

The frequency responses measured from the headphone output and from the microphone output are presented in Figures 4.12 and 4.13, respectively. In Figure 4.12 the solid black curve is obtained from the left channel (active channel) of the headphone output and the red line represents the leakage from the right channel while the equalization was set on. The black dashed curve illustrates the frequency response measured from the left (active) headphone output while the equalization was set off. The excitation signal (black dashed curve) was similar for the equalization off and the equalization out settings. As can be seen, the cross-channel leakage mainly happens while the equalization is set on. The channel separation, with equalization on, is approximately 30 dB. The sharp peaks, in both Figures, seem to be caused by the 50 Hz AC current hum. The highest peak is located at 50 Hz and the rest of the peaks are its odd harmonic components.

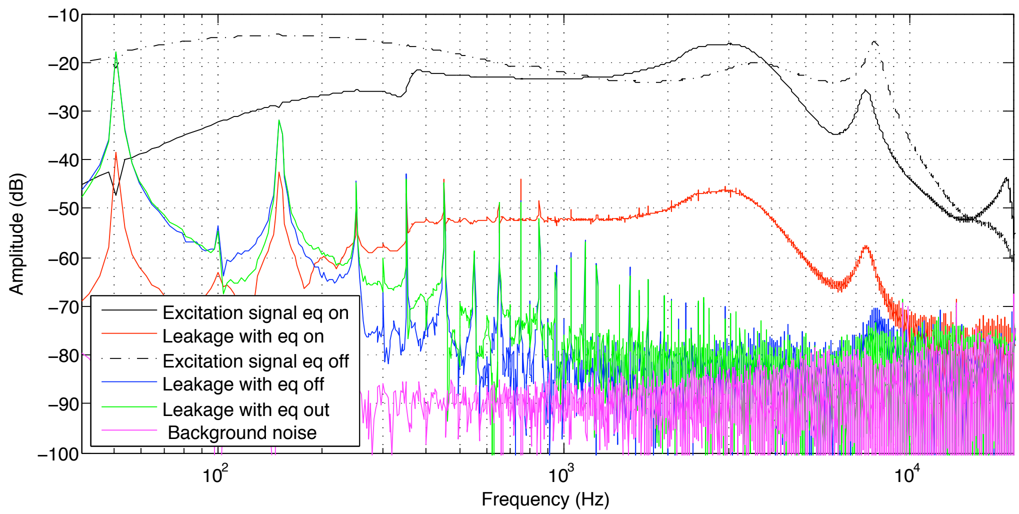

Figure 4.12: Channel separation of the ARA mixer measured from the headphone output.

The channel separation measured from the microphone output (see Figure 4.13) was approximately 30 dB at high frequencies (over 7 kHz) and 30 - 40 dB at low frequencies depending whether the equalization was set on or off. In theory, the equalization settings should not affect the measurement results, because the signal obtained from the microphone output has not been equalized (see Figure 4.11). From a psychoacoustic point of view a channel separation of 25 - 30 dB or better is sufficient for headphone reproduction.

4.6 Other Measurements

4.6.1 Passive Isolation of the Philips Earphone

One advantage of the in-ear headphones is that they passively isolate the external ambient noise. A good fitting of the earphone is extremely important, because of the leakages that happen between the earphone cushion and skin (see Section 3.2.3). The isolation is highly

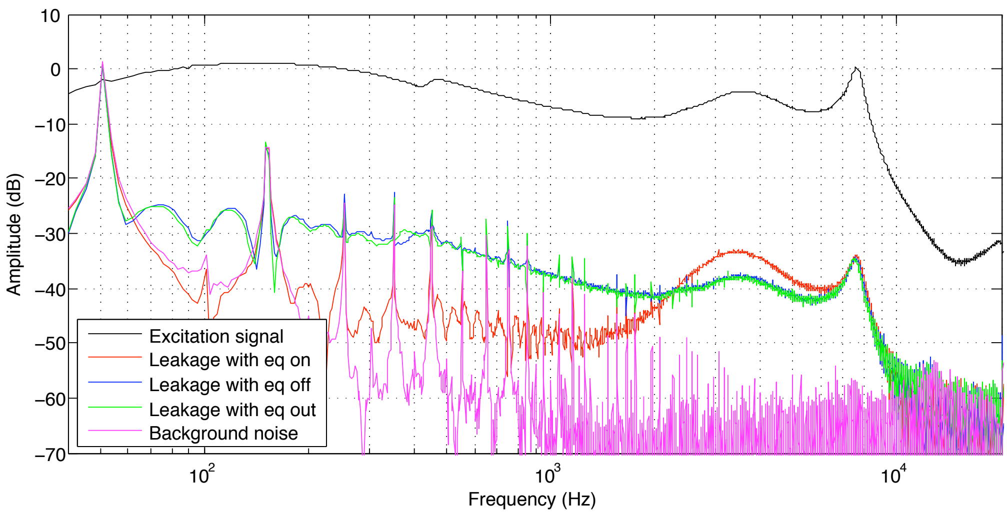

Figure 4.13: Channel separation of the ARA mixer measured from the microphone output.

frequency dependent, attenuation happens only at mid and high frequencies.

The isolation measurements were conducted in the large anechoic chamber at the Department of Signal Processing and Acoustics in Helsinki University of Technology. Two subjects were used in order to obtain a natural fitting for the earphone. A Genelec 8030A loudspeaker was placed in front of the test subject, 3 meters apart. An Edirol FA-101 firewire audio interface was used during the measurements. The measurements were done with Matlab software. A logarithmic sweep was played from the Genelec loudspeaker and the impulse response was recorded inside the open ear canal, with a Knowles FG-3329 miniature microphone. The Knowles microphone was inserted approximately two centimeters deep into the ear canal. Next, the ARA headset was inserted into the ear of the test subject, while the Knowles microphone remained as stationary as possible inside the ear canal. The logarithmic sweep was then played again from the Genelec loudspeaker and recorded inside the ear canal while the ear was occluded with an ARA headset. Figure 4.14 shows a frequency response of an open ear canal and a frequency response of an occluded ear canal. Figure 4.14 demonstrates that all frequencies above 400 Hz are attenuated, however, the ARA headset does not attenuate frequencies under 400 Hz at all. The cord of the microphone was between the skin and the earphone cushion and could have caused some leakage.

A single curve that represents the isolation of an earphone very well can be obtained by taking the dB scale difference between an open ear canal measurement and an occluded ear canal measurement. The isolation curve derived from Figure 4.14 is shown in Figure 4.15.

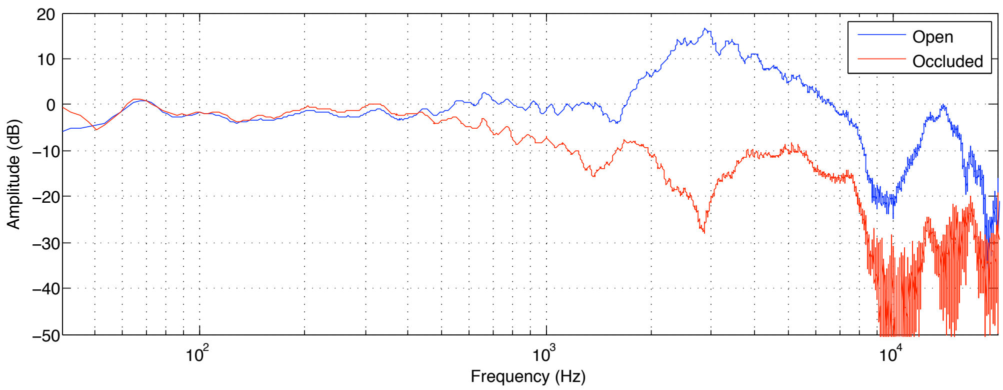

Figure 4.14: Frequency response measured from inside of an open ear canal (blue) and frequency response measured from inside of an occluded ear canal (red).

The 0 dB flat line means that there is no attenuation and the below zero values represent the amount of attenuation.

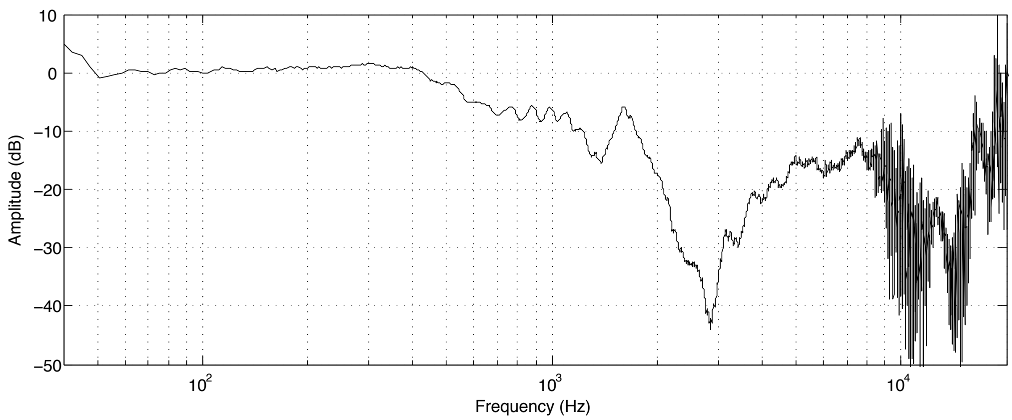

Figure 4.15: An isolation curve of the Philips SHN2500 headphones.

4.6.2 Impact of the Ear Canal's Lenght and the Position of the In-Ear Microphone

The impact of the length of the ear canal and the position of the in-ear microphone can be easily measured with the help of the adjustable ear canal simulator, ADECS (see Section 4.1.2). The length of the ear canal affects the ear canal resonances according to the closed tube Equation (3.1).

The Philips earphone with mounted in-ear microphone (see Section 4.1.5) was inserted into the ADECS. A logarithmic sweep was played through the Philips earphone and the impulse response was recorded with ADECS from the "ear drum". The used lengths of the ADECS ear canal were between 20 - 30 mm at intervals of 1 mm. The Philips earphone goes about 5 mm deep into the ADECS tube, thus the actual lengths of the ear canal were approximately between 15 - 25 mm. Figure 4.16 demonstrates the impact of the length of the ear canal. As can be seen, the half-wave resonance travels from 9.8 kHz to 6.7 kHz.

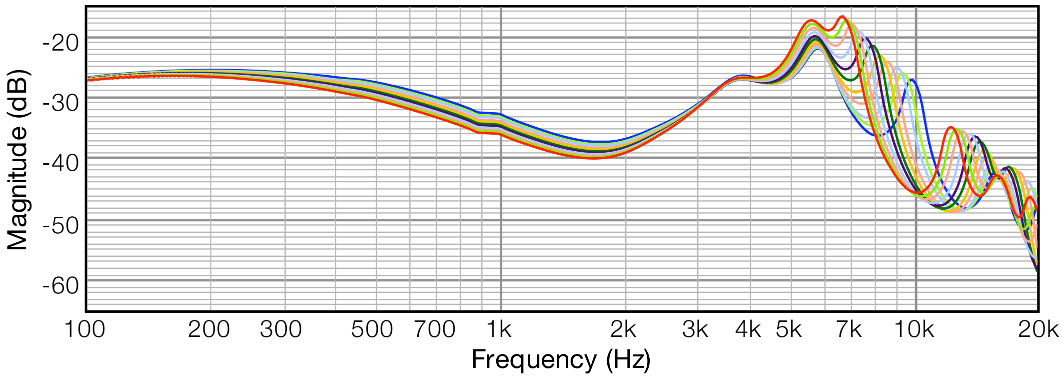

Figure 4.16: Frequency responses of different lengths of the ear canal measured with ADECS. The lengths of the ear canal are between 20 - 30 mm at intervals of 1 mm, the blue curve is 20 mm and the red curve is 30 mm.

The position of the in-ear microphone also affects the ear canal measurements. Figure 4.17 shows five frequency responses, which are measured with ADECS microphone at different distances from the earphone, inside the ear canal. As can be seen, the frequency responses are quite dissimilar. The resonance peaks are at the same frequency, but their magnitudes differ. The major differences between the curves are locations and magnitudes of the anti-resonances (notches).

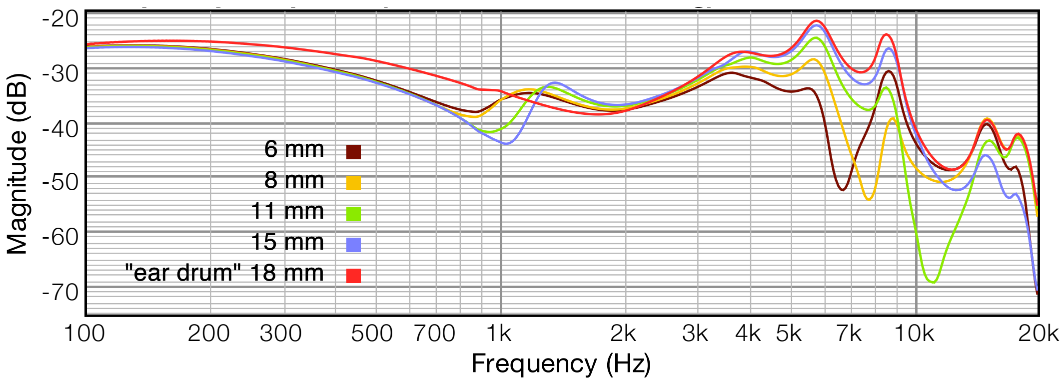

Figure 4.17: The effect of the position of the in-ear microphone measured with ADECS and Philips SHN2500 earphone. The in-ear microphone was placed at five different distances from the earphone.

Chapter 5

Usability of an Augmented Reality Audio Headset

Although many people are using in-ear headphones daily in everyday-life situations, there are still very few studies on how in-ear headphones (in this case ARA headset) would be perceived and accepted when worn for longer periods of time in everyday-life situations. A preliminary study has been conducted in [37]. Based on the preliminary study the overall usability and sound quality was found adequate in most situations. The preliminary study also indicated that without any applications the ARA headset does not benefit the user enough to motivate the usage of the ARA headset for long periods.

One very important factor in the acceptability of the ARA headset is sound quality. The sound quality should be good enough that people can wear the ARA headset for long periods in everyday-life situations without being disturbed by the sound quality of the headset. Thus, the pseudoacoustic representation should be as similar to the real sound environment as possible. Therefore, a proper equalization of the headphone is a vital part of the sound quality of the ARA headset (see Section 3.3.2).

Spatial hearing is another factor that plays an important role in the acceptability of the ARA headset and it should be preserved as normal as possible. The ARA headset should provide the same spatial cues as does the normal hearing situation. The closer the microphones are to the ear canals the more natural are the spatial cues as well [9].

Furthermore, the ARA hardware portability affects the usability. Using the ARA headset should require only little additional equipment and it should be wearable. People are fairly used to carry electronic devices, such as cellular phones and mp3-players, with them in everyday basis and they are even used to wear headphones for rather long periods of time. Thus, carrying the ARA mixer and wearing the ARA headset should not be a problem.

Although the ARA hardware seems to operate quite well based on the preliminary study, building an ideal ARA headset and mixer is very challenging. Some alterations compared to normal hearing situation are known to occur. The earphones fit tightly into the ear canal entrance causing the occlusion effect (see Section 3.2.4). The binaural microphones remain approximately 1 cm from the ear canal entrance, thus, possibly degrading the spatial hearing compared to the normal hearing situation. The microphones have also some directivity at high frequencies, as demonstrated in Section 4.4.3.

5.1 Methods

The usability evaluation was planned and performed by Miikka Tikander [35]. The usability of the ARA headset was tested in everyday-life situations. The usability evaluation consisted of two parts; the actual field test, where the test subjects wore the headset; and a sound quality and adaptation test, where the overall sound quality and usability was tested before and after the field test period. All of the ARA mixers used in the evaluation were generically equalized (see Figure 3.10).

Test Subjects

A group of ten testees (9 male, 1 female) took part in the usability evaluation. No one reported having any kind of hearing related problems. All testees had technical background and some experience in evaluating sound quality. The test group had also a positive attitude towards technical appliances, which was important because of the size and impracticality of the current ARA mixer prototype. Furthermore, it was important that the testees were willing to wear the headset in all kinds of situations, even when it might have been slightly impractical or uncomfortable. It should be taken into account that the ARA headset is designed to be used by ordinary people. However, the prototype character of the ARA mixer could have drawn too much attention away from the actual test, if a set of technically unexperienced testees would have participated it.

Field Test

The field test part of the usability evaluation consisted of real-life events. The testees were given an ARA headset and mixer, in order them to wear it in their everyday-life. Testees were instructed to wear the ARA headset for long periods of time in as many everyday-life situations as possible. They were even encouraged to find all kinds of new approaches to use the device. The testees wrote all their observations down in a diary. A guiding question for the testees to keep in mind during the test was, "Would you be willing to wear this kind of a headset in real-life situations?".

Between the testees, the evaluation period lasted from four days to a week, and the total usage of the headset varied between 20 - 40 hours. The continuous usage periods, during the field test, were between few minutes to eight hours.

Controlled Usability Evaluation

A set of controlled usability evaluations were conducted for the testees, before and after the field test. Before the first usability test, the testees had some time to familiarize themselves with the ARA headset and get used to wearing it. The usability test consisted of four tasks: Reading text out loud, eating crisp bread, drinking water, and chewing bubble gum. All the tests related to the annoyance of the occlusion effect.

In the text reading part, the testee was given a sheet of text which he/she had to read out loud paragraph by paragraph. The paragraphs were read twice, first with the headset on and then without the headset. The test subjects were informed that they could read the texts as many times as it was necessary to form an opinion on the annoyance of the ARA headset in this task. In crisp bread eating, water drinking and bubble gum chewing the testees were similarly evaluating the annoyance of the activity with the headset on compared to the normal case without the headset. The testees were instructed to grade the annoyance as if they had to perform the activities in everyday life situations.

The annoyance was evaluated on a standard MOS-scale (Mean Opinion Score) [10]. The test subject evaluates the annoyance level of the sample compared to a reference sample, that is, the annoyance of the activity with the headset on compared to the normal hearing experience. The grade was done with a pen on a paper by marking the grade on a continuous scale. The MOS-scale is shown in Table 5.1. The scale goes from one to five, where five is the best and one is the worst.

Table 5.1: MOS grading scale.

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

Sound Quality

The sound quality of the ARA headset was evaluated with natural sounds, such as speech and finger snaps, as well as with two stereo playbacks. The sound quality evaluation was made in a standardized listening room. The test group was asked to evaluate the following attributes: spatial impression, timbre, and location accuracy, with the same MOS-scale that was used in the controlled usability evaluations. Spatial impression was described to include the overall sense of space, especially room size and reverberation. Timbre was instructed to denote tonal coloration, and location accuracy was described to measure spatial hearing, including the direction and distance of the sound source.

The sound quality evaluation started with the human voice and finger snapping. The tester spoke out loud and snapped his fingers in different locations of the room. The testees were instructed to listen to the tester with and without the ARA headset. The testee was also allowed to ask the tester to move to a specific location, if the testee felt like he/she needed more information. The predetermined locations of the tester are depicted with small circles in Figure 5.1. After this part, the testees listened to two stereo music samples. The testees were allowed to listen to the music samples as many times as they thought was necessary in order them to form an opinion. Finally, the testees were asked to evaluate the sound quality based on all of the above-mentioned samples.

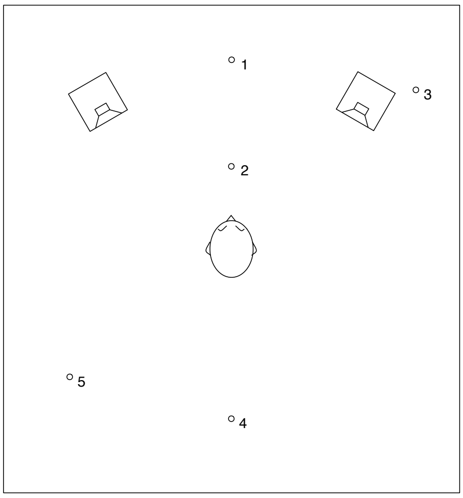

Figure 5.1: Listening test setup.

Adaptation Evaluation After the Field Test Period

The controlled usability and sound quality evaluations were repeated after the field tests. The testees were instructed to wear the ARA headset continuously at least 1.5 hours before the evaluation. This way the testees had adjusted to the ARA headset and the pseudoacoustic environment and estranged from the natural acoustic reference, at least to some extent. First, the evaluation was done with no comparison of any kind, that is, the testees did not remove the headset during the test. After that, the evaluation was repeated similarly as before the field test, that is, the testees compared the pseudoacoustic environment, with the headset on, to the normal hearing situation, without the headset. The grading was done the same way as before.

Post Evaluation Interview

After the final evaluations, the testees were interviewed. During the interview sessions, the diaries were reviewed and discussed. Furthermore, the testees were asked to evaluate some common usability issues numerically, using the same MOS-scale as before. The usability issues evaluated were; cords (usability and mechanical noise); Annoyance of artifacts in general, such as distortion, occasional clippings, pops, and radio interference; using mobile phone (handset on the ear); and walking.

5.2 Results

5.2.1 Usability and Adaptation Evaluation

Figure 5.2 shows the results of the controlled usability and sound quality evaluations. Each attribute has three bars corresponding to the evaluation before the field test period (left-most bar), the evaluation after the field test without reference to normal hearing situation (middle bar), and the evaluation after the field test with reference (right-most bar). Median for each case is denoted by a horizontal line inside the boxes. Upper and lower ends of the boxes correspond to upper and lower quartile values (the median of the upper and lower half of the data). The whiskers extending from each end of the boxes show the smallest and the largest value of the data set that is not an outlier. Furthermore, the plus signs outside the boxes are data points considered to be outliers and the plus signs inside the boxes denote the mean value of each data set.

Figure 5.2 suggest that the most annoying activities were crisp bread eating and bubble gum chewing, whereas the water drinking was found the least annoying. As can be seen, the results were not entirely consistent within the test group, thus the perceived annoyance of the activities seems to be highly individual. On the other hand, the sound quality evaluations indicated that the overall quality of the pseudoacoustic representation was found rather good. Especially, the spatial impression and location accuracy were perceived exceptionally well.

Figure 5.3 shows the intra-subject difference between the pre- and post-field-test evaluation of the usability and sound quality. The left-most bar in each attribute corresponds to the difference between the evaluation before the test and the evaluation after the test

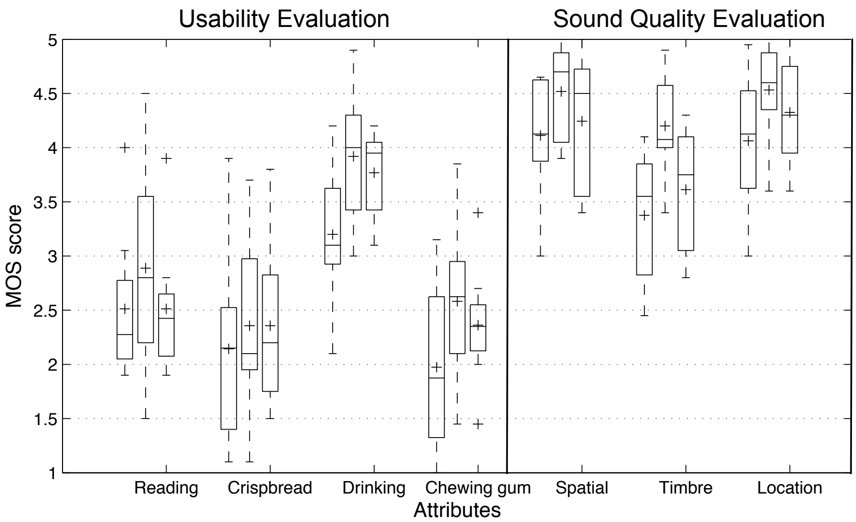

Figure 5.2: Usability and sound quality evaluation results. Left-most bar in each triplet corresponds to evaluation before the field test period, middle and right-most bars correspond to evaluation after the field test period without reference to open ear, and with reference to open ear, respectively.

without reference to natural hearing situation (open ear). The right-most bars correspond to the difference between the evaluation before the test and the evaluation after the test with reference. The positive values correspond to adaptation, that is, the testee has evaluated an attribute to be less annoying after the field test period compared to the evaluation conducted before the field test period.

As can be seen, Figure 5.3 indicates that the adaptation concerning the usability evaluation is quite low. However, the adaptation to the sound quality seemed to be more consistent, especially when there were no reference to the natural hearing situation (the left-most bars). Even with reference to the natural hearing situation, the testees evaluated the sound quality to be better after the field test. One possible reason for the low adaptation for the reading and eating related usability evaluation might be that the internal sounds due to these activities are so well known that the 1.5 hours is not enough to fully forget the feeling or sense of these activities. Furthermore, the occlusion effect is quite loud, thus it makes the adaptation more difficult.

Based on the interviews of the test subjects during the evaluations, the crisp bread eating was found quite noisy, in general, because the occlusion effect enhances the loudness of every bite. The bubble gum eating was not considered as loud as the crisp bread eating, but the fact that the chewing is going to take some time led the test subjects to grade the activity

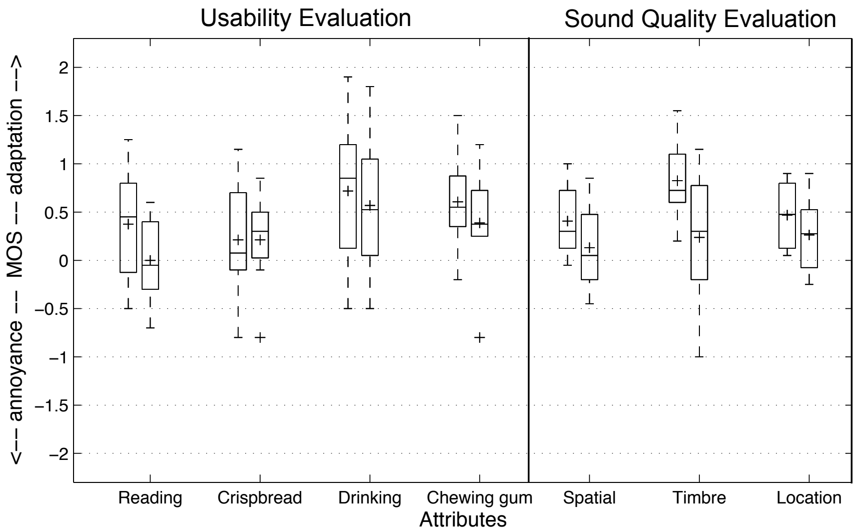

Figure 5.3: Intra-subject difference between the usability and sound quality evaluations before and after the field test.

more annoying. Similarly, the drinking was found mostly not annoying because of the short duration of the event.

The results of the interviews after the field test period are shown in Figure 5.4. In overall, the cords were found very annoying and impractical. The cords conducted mechanical noise to the headphones, which was found very irritating. Furthermore, the cords tend to get stuck and mixed in clothes and in other objects. In fact, one of the main reasons why the awareness of the headphones was present all the time was because the user heard all the movements of the cords.

Occasional artifacts such as clipping due to loud bangs, radio interference, and pops when turning switches were not perceived annoying. However, using a mobile phone, holding the phone normally on the ear, was found very annoying. The phone was hard to couple with the pinna because the headset extended slightly outside the ear canal. This resulted in very thin and low sound and it also caused annoying mechanical noise, when the mobile phone touched the headset. In general, walking was found fairly normal. However, this only applies to just walking, it does not consider any corollary annoyance factors such as mechanical noise from the cords or conversation while walking.

5.2.2 Interviews and Diaries

This section discusses and summarizes the interviews and diaries made by the testees. The overall feedback has been very positive. The sound quality was experienced good and

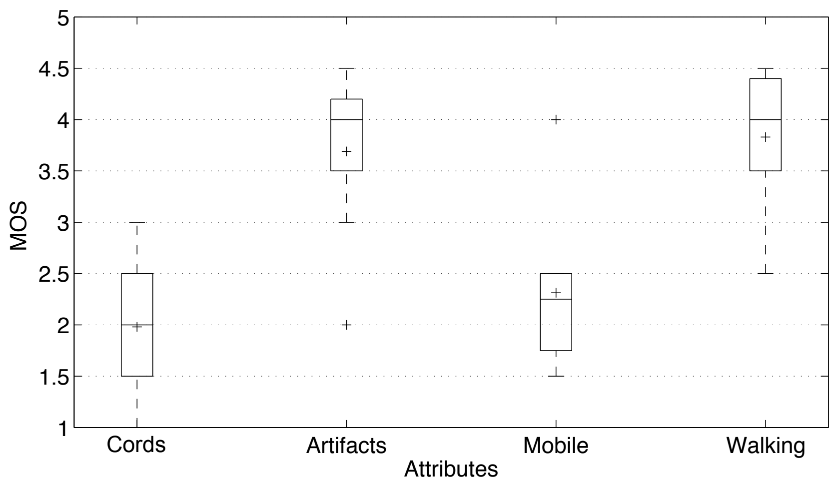

Figure 5.4: MOS values given in the interview after the field test.

applicable for the most part of normal everyday-life situations. The most discussed attribute was the cords. As mentioned before, the cords tend to get stuck in clothing and create mechanical noise to the headset.

Sound Quality

The sound quality of the headset was found very good. Spatial and directional hearing was experienced to be very close to the natural hearing situation. The sound color (timbre) was found sufficiently good for everyday use. Many testees complained of slightly attenuated high frequencies. The attenuation could be perceived when listening to familiar wide-band sounds such as air-conditioner, running water or swishing leaves. Although this was found perceptible, it was not reported to be annoying.

All of the testees reported that sound sources were very well externalized, even sound sources located in front of the testee. The reason for very good externalization in front is probably the synergy between the hearing and the vision, as it happens in real-life situations. When binaural signals from the headset were recorded and then listened to later on, the frontal externalization was not that good anymore. This is understandable because the visual cues and the head movement cues are no longer present in the recordings. Another reason could also be the channel separation of the ARA mixer, see Section 4.5.1. One testee noted that sometimes non-visible and unfamiliar sounds where located inside the head.

The electret microphones of the ARA headset created some audible noise (approximately A-weighted level of 27 dB, see Section 4.4.2). The testees reported that the noise was noticeable at first, but they got used to it quite fast and after some time no attention was paid to it. The noise might slightly raise the hearing threshold, however, the noise was not reported to be annoying. One often heard comment was that when wearing the ARA headset the testees paid more attention to the surrounding sounds; for example, some familiar sounds might have sounded strange, but when the headset was removed the sound still sounded the same. In that sense, the testees expected some alterations in the pseudoacoustic representation, even when there were none. Some testees, quite the contrary, reported that they had missed their mobile phone ringing because the ringing tone sounded different with the ARA headset on.

Ergonomics

The handling noise and limitations caused by the cords of the headset was reported annoying by every test subject. The cords tend to get stuck in clothes and limit the head movement. A few testees reported that they could get used to the headset itself and forgot they are wearing it, but the mechanically transmitted noise from the cords keeps reminding about the headset. This was found very annoying as well. Some wireless technique, such as Bluetooth, could be used between the mixer and headset in order to eliminate the impracticalities of the wires. Another issue was the size of the ARA mixer. It is the first prototype of the mixer and it is, in fact, rather large. The ARA mixer was reported inconvenient to carry, because it does not fit in regular size pockets. Technically, the size of the ARA mixer could be easily made smaller. The prototype of the ARA mixer has one-sided circuit board, which could be easily fitted in a smaller (e.g, two-sided) circuit, thus solving the problem.

Placing and removing the headset was found trouble-free. However, majority of the testees commented that they had some ear ache in the beginning of the field test period. The ear ache emerged typically after 1.5 hours of usage at the first day of the field test period. Nonetheless, after the first day the testees accustomed to the headset and no further ear ache was reported. A few test subjects reported that they had ear ache throughout the test period and therefore had troubles wearing the headset any longer than 20 - 50 minutes at a time. Furthermore, some test subjects reported that after a certain period of the headset usage their ear started to itch, whereas some reported they they had itchy feeling after a long period of use.

The headset extended slightly outside the ear canal, which had detrimental effect on using a mobile phone. The whole test group reported that using a mobile phone was cumbersome. When the mobile phone touched the ARA headset it created an unpleasant contact noise in the headset. Furthermore, the coupling of the mobile phone and the headset was bad, thus resulting an indistinct sound quality. One solution to this problem could be a headset that extends less from the ear canal entrance. However, the best solution would probably be to implement a hands-free functionality to the ARA mixer and use the headset as a hands-free device for the mobile phone. In fact, preliminary tests have shown that using the headset as a hands-free device is very practical and useful.

Communications

All test subjects reported that the sound quality of speech was good and that they had no problems in normal conversation situations. One big problem reported was the sound of user's own voice; it sounded boomy and hollow due to the occlusion effect. In addition, the testee's own voice was localized inside the head. However, most of the testees habituated to this quite fast and towards the end of the test period they found it less annoying. Some testees had sometimes troubles to determine the level of their own voice. A commonly complained issue was that conversations were troublesome during dinners and while walking. The occlusion effect enhances the eating sounds so much that they mask the sounds from the conversion. Similarly, while walking, the moving cords inflict mechanical noise to the headset and the foot steps produce thumbing sound to the ear canals due to the occlusion effect. This was found relatively annoying and it was mentioned to be somewhat exhausting as it required more mental focusing for listening.

Most of the testees reported to have felt social discomfort when wearing the headset in places where it was not necessarily appropriate (e.g., in lectures or conversations). A few testees commented that they felt they had to somehow prove that they are, in fact, paying an attention; for example, by answering to a question faster than normally during a conversation. This problem would not exist if people knew that the headset is hear-through, thus enabling normal communication. Therefore, this problem will probably decline when the time elapses and the ARA headsets become more common.

Other Observations

Loud sounds, such as bangs from closing doors, overloaded either the microphones or the amplifiers resulting in distorted sound in the ARA headset. Some of the testees found this very irritating, while the others did not mind it that much. Continuous loud sounds (e.g, rock concert) resulted in distorted sound as well. The distortion could be prevented by, e.g, implementing better electric design. Furthermore, some of the connectors of the cords caused undesired scratchy noises, when there were contact problems due to the movement of the cords.

Although the testees reported to get accustomed to the thumbing sounds of walking fairly fast, running and jumping was still reported to be unpleasant. Furthermore, heavy breathing after an exercise sounded boosted and was also reported unpleasant. Also, strong wind caused displeasing sounds in the headset, whereas normal wind sounded quite natural.

Some of the testees listened to music via the ARA mixer. The ARA mixer mixes the music and the natural sound surroundings. It was commented that it was nice to hear the full frequency range of the music, as well as the surrounding sounds. This worked well in a quiet environment, however, the downside was that the music level had to be risen to fairly high, in order to hear the music in a noisy environment (e.g., in a bus). Furthermore, the music was localized inside the head, just like with normal headphones. This was reported to be an advantage as it helped the separation of the music and the real surrounding sounds.

5.2.3 Conclusions and Feature Suggestions

The usability evaluation indicates that a generically equalized ARA headset is applicable in everyday-life situations. The sound quality and usability was found sufficiently good for long-term usage. However, the ARA headset as such does not bring any additional value to listening to surrounding sounds. Thus, if the ARA headset would have some extra advantage, such as ARA applications (see Section 3.5), the test group would be willing to use the headset, even for long periods at a time. Furthermore, the results indicate that the design of the ARA mixer should be improved. Yet, the most desired feature was the wireless headset, which would, indeed, increase the usability of the ARA headset considerably.

Another feature suggestion reported by many testees was the ability to adjust the level of the pseudoacoustic sounds. This way the user could, e.g., attenuate the pseudoacoustics when listening to music in a noisy environment or enhance it when having a conversation. Mobile phone hands-free functionality was also a very desired feature.

Chapter 6

Conclusions and Future Work

6.1 Conclusions

Detailed laboratory measurements and a full-scale usability evaluation of the ARA hardware was conducted. The results of the usability evaluation were consistent with the preliminary evaluation [30]. The results suggested that the sound quality and usability of the ARA headset are sufficient for everyday usage. However, the large size of the ARA mixer was found slightly impractical. Furthermore, the ARA headset itself does not offer any enhancement to the surrounding sounds, thus ARA applications are needed.

The laboratory measurements did not expose any major deficiencies related to the ARA hardware. Distortion of the earphone drivers was barely audible and better than average when compared to other in-ear headphones (see Figure B.2). The self-noise of the ARA headset was found audible in quiet environments based on both the usability evaluation and the laboratory measurements. However, it was not reported to be annoying. The directivity of the headset microphones was also rather insignificant. The channel separation of the ARA mixer was found to be sufficient for headphone reproduction.

6.2 Future Work

There are plans to develop an advanced 2.0 version of the ARA mixer. Although the usability of the ARA headset was found sufficient in most everyday-life situations, there is still need for many upgrades. The major problem with the current ARA mixer is the size, thus the main objective is to reduce the size of the ARA mixer. Normal mobile phone usage is also troublesome, so an integrated hands-free functionality to the ARA mixer would improve the usability considerably. Every test subject reported that the headset cords were really cumbersome, thus the headset should be implemented wirelessly. The occlusion effect, caused by the in-ear headphones, is a big problem. It is a difficult problem to overcome but, e.g., an active occlusion cancellation could be implemented to the ARA mixer. Furthermore, one very desired functionality was the ability to adjust the sound pressure level of the pseudoacoustic environment.

A better generic equalization could be obtained by measuring a larger group of people. Furthermore, it would be interesting to research how different individually made equalizations affect to the sound quality. In the future we plan to conduct a listening test where different equalizations are compared to the normal hearing situation. These equalization options could be: current generic equalization, improved generic equalization, individual equalization with the ARA mixer, individual equalization with graphic equalizer and even individual equalization with digital signal processing.

Other important additions for the ARA technology that need to be developed are the ARA applications. As mentioned before, the ARA headset needs these applications before it starts to bring additional value to everyday usage. All in all, the next big thing that ARA technology needs is a killer application.

Bibliography

[1] Audacity software, http://audacity.sourceforge.net/.

[2] http://hyperphysics.phy-astr.gsu.edu/hbase/sound/ear.htmlc1.

[3] Matlab software, http://www.mathworks.com/products/matlab/.

[4] Playrec multi-channel matlab audio sofware, http://www.playrec.co.uk/.

[5] T. Beard. Personal hearing control system and method. United States Patent Application Publication, May 2008.

[6] B&K. *Product Data, Head and Torso Simulator - Type 4128C*.

[7] J. Blauert. *Spatial Hearing*. The MIT Press, Cambridge, Massachusetts, 1983.

[8] J. Borwick. *Loudspeaker and Headphone Handbook*. Focal Press, second edition, 1994.

[9] D. Brungart, A. Kordik, C. Eades, and B. Simpson. The effect of microphone placement on localization accuracy with electronic pass-through earplugs. *IEEE Workshop on Applications of Signal Processing to Audio and Acoustics*, October 2003.

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

[11] N. I. Durlach and H. S. Colburn. Binaural phenomena. *Handbook of perception: Hearing*, 4:365–455, 1978.

[12] J. Eargle. *The Microphone Book*. Focal Press, 2nd edition, 2004.

[13] A. Farina. Simultaneous measurement of impulse response and distortion with a swept-sine technique. *AES 108th Convention, Paris*, February 19-22 2000.

[14] D. Hammershøi. Fundamental aspects of the binaural recording and synthesis techniques. *100th AES Convention*, May 1996.

[15] 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.

[16] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, H. Nironen, and S. Vesa. Techniques and applications of wearable augmented reality audio. *J. Audio Eng. Soc.*, 51:419, May 2003.

[17] M. Hiipakka. Measurement apparatus and modelling techniques of ear canal acoustics. Master's thesis, Helsinki University of Technology, December 2008.

[18] M. Karjalainen. Augmented reality for audio applications, project report. Helsinki University of Technology, Finland, Department of Signal Processing and Acoustics.

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

[20] J. W. Lee and J. M. Lee. Forced vibro-acoustical analysis for a theoretical model of a passenger compartment with a trunk - part 1: Theoretical part. *Journal of Sound and Vibration*, (299):900–917, 2007.

[21] N. Lenihan. Wlan positioning. University of Limerick, Wireless Access Research.

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

[23] J. Mejia, H. Dillon, and M. Fisher. Active cancellation of occlusion: An electronic vent for hearing aids and hearing protectors. *J. Acoust. Soc. Am.*, 124(1):253–240, July 2008.

[24] S. Mitra. *Digital Signal Processing, A Computer-Based Approach*. McGraw Hill Higher Education, 2nd edition, 2001.

[25] 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.

[26] Nokia. Fashion stereo headset hs-3, http://europe.nokia.com/a4471219.

[27] Nokia. Pop-port^{TM}, http://www.maps.nokia.com.

[28] Nokia. *Nokia AV Hardware Interface Specification*, 2008.

[29] C. Poldy. Headphone fundamentals. *AES 120th, Paris*, 2006.

[30] V. Riikonen. User-related acoustics in a two-way augmented reality audio system. Master's thesis, Helsinki University of Technology, April 2008.

[31] V. Riikonen, M. Tikander, and M. Karjalainen. An augmented reality audio mixer and equalizer. *124th AES Convention*, May 2008.

[32] T. Rossing, P. Wheeler, and R. Moore. *The Science Of Sound*. Addison Wesley, third edition, 2002.

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

[34] M. Tikander. Modeling the attenuation of a loosely-fit insert headphone for augmented reality audio. *AES 30th International Conference*, March 2007.

[35] M. Tikander. Usability of an augmented reality audio headset. *Submitted to the Journal of the Audio Engineering Society (JAES)*, 2008.

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

[37] M. Tikander, M. Karjalainen, and V. Riikonen. An augmented reality audio headset. *11th Int. Conference on Digital Audio Effects (DAFx)*, September 2008.

[38] A. Walker, S. Brewster, D. McGookin, and A. Ng. Diary in the sky: A spatial audio display for a mobile calendar. Glasgow Interactive Systems Group, Department of Computing, 2001.

[39] C. Youngblut, R. Johnson, S. Nash, R. Wienclaw, and C. Will. Review of virtual environment interface technology. Institute for Defense Analyses, 1996.

Appendix A

The ARA Mixer

The ARA mixer has ten external connectors and controllers; six 3.5 mm stereo-mini plugs, two slide switches and two adjustable controls. The external connectors and controllers are depicted in Figure A.1. There are also 14 adjustable controls inside the ARA mixer. These controls determine the equalization properties for both channels (7 controls per channel).

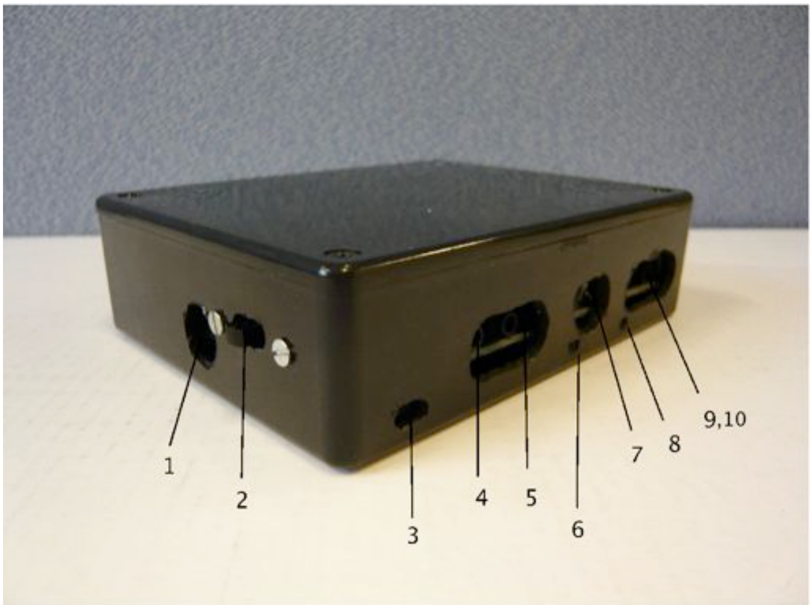

Figure A.1: The ARA mixer's externals connectors and controllers [30].

1. Output for an external equalization device
2. Power switch

- Left position: Power off
- Right position: Power on

3. Equalization switch
   - Left position: External equalization connector (1) activated
   - Middle position: Equalization off
   - Right position: Equalization on
4. Unequalized output
5. Microphone input
6. Level screw for the right channel
7. Headphone output
8. Level screw for the left channel
9. Input 1
10. Input 2

The ARA mixer uses two 9 V batteries as a power supply. They should provide approximately 20 hours of use [30]. The equalization only affects the microphone signals routed into user's ears. So the signals from the outputs are unequalized binaural microphone signals. The inputs are also unequalized, the signals from the inputs are routed directly to the headset.

The outputs (1) and (4) are otherwise identical except that the first one is activated with a slide switch (3). Output (1) is meant for an external equalization device. The external equalization device should be connected between output (1) and either input 1 (9) or input 2 (10), since the inputs are identical. The other input is meant for an external device, such as mp3-player, self phone or computer.

Appendix B

Comparison of Headphones

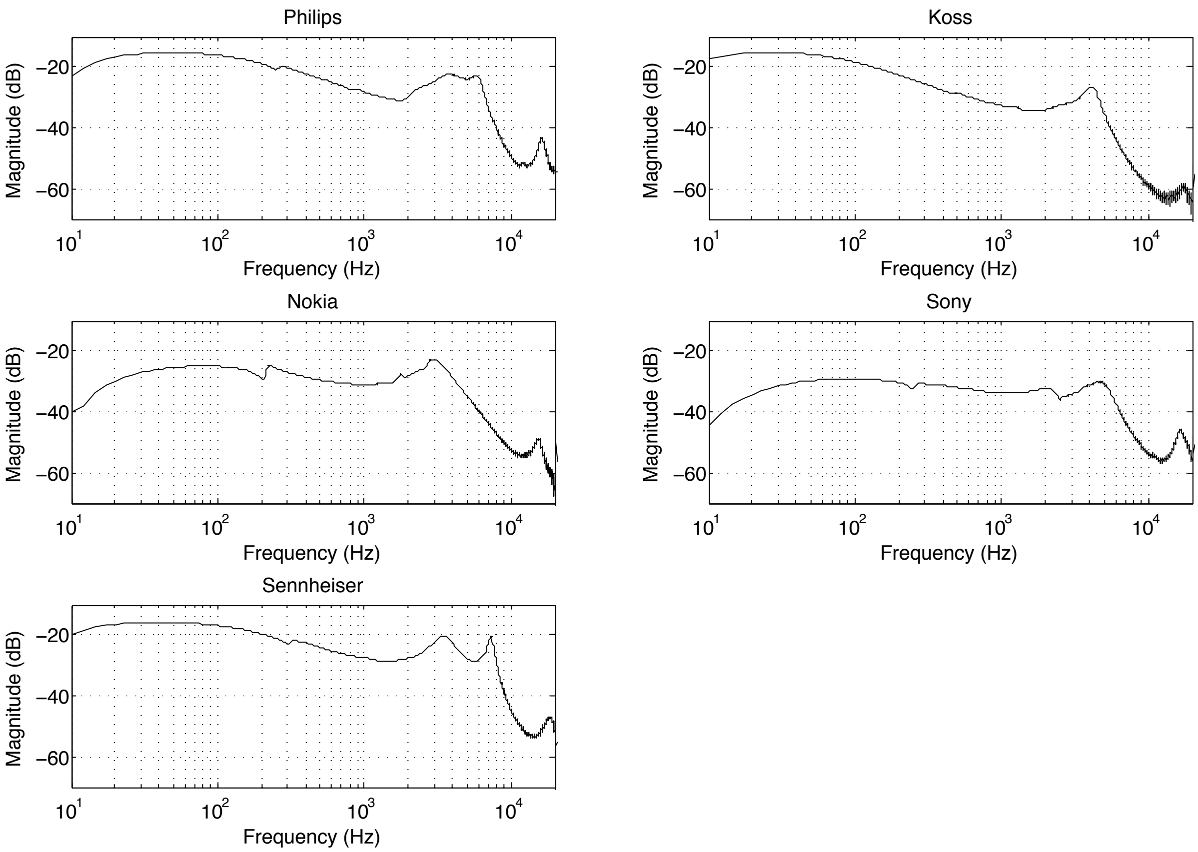

Figure B.1: Frequency responses of Philips SHN2500, Koss KEB24, Nokia in-ear headphones, Sony MDR-EX52 LP and Sennheiser CX300.

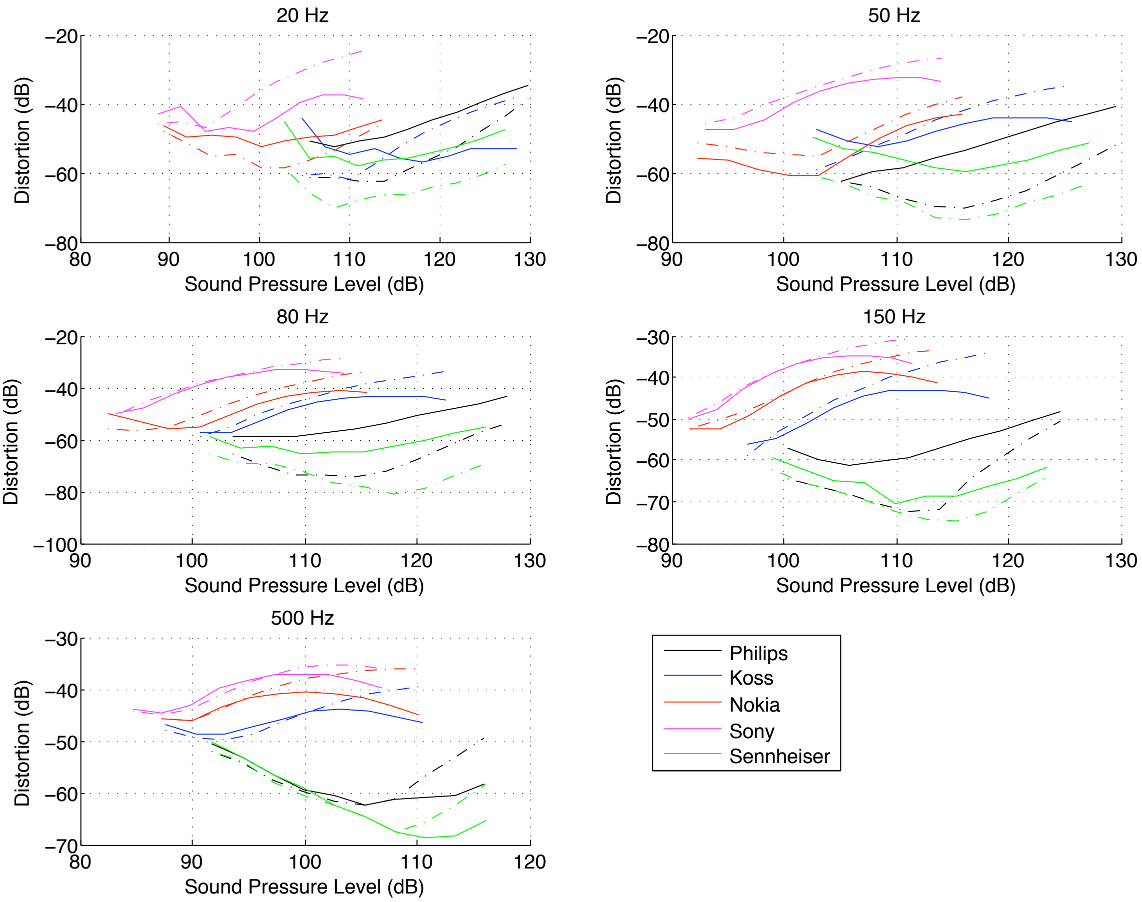

Figure B.2: Distortions of Philips SHN2500, Koss KEB24, Nokia in-ear headphones, Sony MDR-EX52 LP and Sennheiser CX300. The solid lines are the 2^{nd} harmonic components and the dash lines are the 3^{rd} harmonic components.

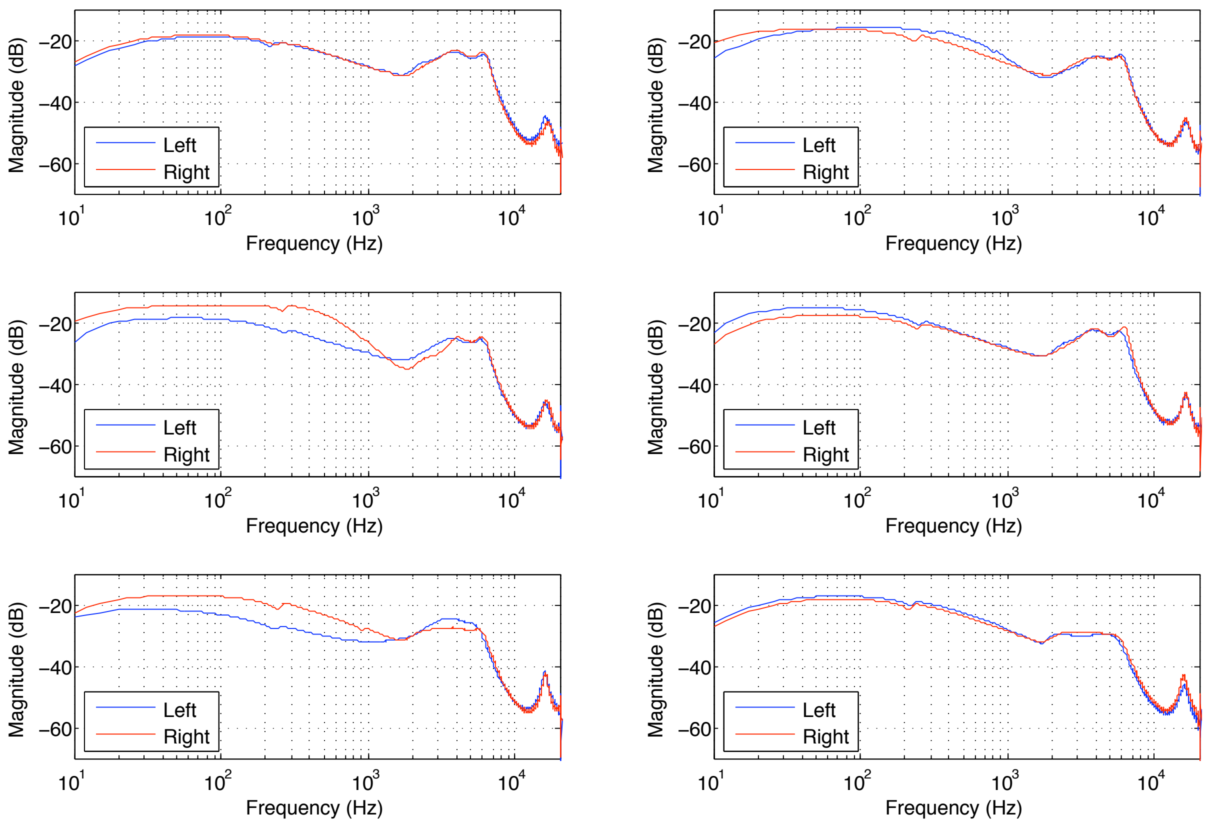

Figure B.3: Frequency responses of different pairs of Philips SHN2500 headphones.