Helsinki University of Technology
Department of Signal Processing and Acoustics
Espoo 2009 Report 13

DEVELOPMENT AND EVALUATION OF AUGMENTED REALITY AUDIO SYSTEMS

Miikka Tikander

Dissertation for the degree of Doctor of Science in Technology to be presented with due permission for public examination and debate in Auditorium S4, Faculty of Electronics, Communications and Automation, Helsinki University of Technology, Espoo, Finland, on the 13th of November 2009, at 12 o'clock noon.

Helsinki University of Technology
Faculty of Electronics, Communications and Automation
Department of Signal Processing and Acoustics

Teknillinen korkeakoulu
Elektroniikan, tietoliikenteen ja automaation tiedekunta
Signaalinkäsittelyn ja akustiikan laitos

Helsinki University of Technology
Department of Signal Processing and Acoustics
P.O. Box 3000
FI-02015 TKK
Tel. +358 9 4511
Fax +358 9 460224
E-mail Lea.Soderman@tkk.fi

ISBN 978-952-248-152-8
ISSN 1797-4267

Multiprint Oy
Espoo, Finland 2009

ABSTRACT OF DOCTORAL DISSERTATION	HELSINKI UNIVERSITY OF TECHNOLOGY P. O. BOX 1000, FI-02015 TKK http://www.tkk.fi

Author Miikka Tikander

Name of the dissertation

Development and evaluation of augmented reality audio systems

Manuscript submitted 10.7.2009	Manuscript revised 27.10.2009
Date of the defence 13.11.2009	
☐ Monograph	☒ Article dissertation (summary + original articles)

Faculty	Faculty of Electronics, Communications and Automation
Department	Department of Signal Processing and Acoustics
Field of research	Acoustics and Audio Signal Processing
Opponent(s)	prof. John Mourjopoulos
Supervisor	prof. Matti Karjalainen
Instructor	

Abstract

This thesis explores different aspects of augmented reality audio (ARA). In ARA applications, the natural sound environment is enriched with the addition of virtual sounds, and their mixture is presented to the listener using headphones. A typical configuration has microphones placed outside of the headphones, and these natural environmental signals are combined with virtual source renderings. This work focuses on the concept of using an insert-type of headset with integrated microphones as a platform for ARA applications and environments.

Subjective and objective measurements were performed to yield optimum equalization for the headset design. Based on the measurement results, an ARA system composed of a headset and mixer unit is proposed. The sound quality and usability of the system is further evaluated in laboratory and real-life situations.

Furthermore, a novel method for acoustic positioning and head tracking is introduced. By placing a set of anchor sound sources in the environment, an ARA headset can be used to determine sound propagation times from the anchor sources to the headset microphones. When the sound source locations are known a priori, this information can be further processed to estimate the user position and orientation. In addition to the positioning method, different sound signaling systems for anchor sound sources are also studied.

Keywords Augmented reality audio, headset, modeling, positioning, head tracking, usability

ISBN (printed) 978-952-248-152-8	ISSN (printed) 1797-4267
ISBN (pdf) 978-952-248-153-5	ISSN (pdf) 1797-4267
Language English	Number of pages 138

Publisher Helsinki University of Technology, Department of Signal Processing and Acoustics

Print distribution Report 13, TKK, Signal Processing and Acoustics

☒ The dissertation can be read at http://lib.tkk.fi/Diss/

VÄITÖSKIRJAN TIIVISTELMÄ	TEKNILLINEN KORKEAKOULU PL 1000, 02015 TKK http://www.tkk.fi
Tekijä Miikka Tikander	
Väitöskirjan nimi Lisätyn äänitodellisuuden systeemien kehittäminen ja evaluointi	
Käsikirjoituksen päivämäärä 10.7.2009	Korjatun käsikirjoituksen päivämäärä 27.10.2009
Väitöstilaisuuden ajankohta 13.11.2009	
☐ Monografia	☒ Yhdistelmäväitöskirja (yhteenveto + erillisartikkelit)

Tiedekunta	Elektroniikan, tietoliikenteen ja automaation tiedekunta
Laitos	Signaalinkäsittelyn ja akustiikan laitos
Tutkimusala	Akustiikka ja äänenkäsittelytekniikka
Vastaväittäjä(t)	prof. John Mourjopoulos
Työn valvoja	prof. Matti Karjalainen
Työn ohjaaja	

Tiivistelmä

Tässä väitöskirjatyössä tutkitaan lisätyn äänitodellisuuden (LÄ) eri sovellusalueita, laitteistoa sekä käytettävyyttä. Lisätyssä äänitodellisuudessa käyttäjän luonnollista ääniympäristöä rikastutetaan virtuaalisilla ääniobjekteilla, jotka voidaan toistaa käyttäjälle erityisillä lisätyn äänitodellisuuden kuulokkeilla. Tyypillisesti LÄ-kuulokkeissa mikrofonit on integroitu kuulokkeisiin, ja toistamalla mikrofonien tallentama luonnollinen ääniympäristö suoraan kuulokkeisiin, yhdessä virtuaalisten ääniobjektien kanssa, voidaan käyttäjälle tuottaa rikastettu versio luonnollisesta ääniympäristöstä. Tämä työ keskittyy erityisesti sovelluksiin, jossa käytetään sovellusalustana inserttityyppisiä kuulokkeita, joihin on integroitu mikrofonit.

Optimaalista kuuloke-ekvalisointia varten tässä työssä on suoritettu sekä subjektiivisia että objektiivisa mittauksia. Mittaustulosten perusteella esitetään LÄ-sovelluksille alustaratkaisu, joka koostuu kuulokkeista ja reaaliaikamikseristä. Kuuloke-mikseri -systeemin käytettävyyttä ja äänenlaatua tutkittiin laboratorio- ja kenttäolosuhteissa. Kenttätesti suoritettiin käyttäjän normaaleissa arkiympäristöissä.

Tässä työssä esitetään myös uusia menetelmiä käyttäjän akustiseen paikannukseen ja orientaation estimoimiseen. Sijoittamalla äänilähteitä käyttäjän ympäristöön voidaan kuulokesysteemin avulla mitata äänisignaalin kulkuajat äänilähteistä kuulokkeiden mikrofoneihin. Kun äänilähteiden paikat tiedetään etukäteen, voidaan tästä tiedosta estimoida käyttäjän koordinaatit ja orientaatio. Paikannusmenetelmän lisäksi esitellään erilaisia signalointitapoja kulkuajan mittaamiseen.

Asiasanat Lisätty äänitodellisuus, kuulokkeet, mallintaminen, paikannus, päänseuranta, käytettävyys	
ISBN (painettu) 978-952-248-152-8	ISSN (painettu) 1797-4267
ISBN (pdf) 978-952-248-153-5	ISSN (pdf) 1797-4267
Kieli English	Sivumäärä 138
Julkaisija Teknillinen korkeakoulu, Signaalinkäsittelyn ja akustiikan laitos	
Painetun väitöskirjan jakelu Report 13, TKK, Signaalinkäsittelyn ja akustiikan laitos	
☒ Luettavissa verkossa osoitteessa http://lib.tkk.fi/Diss/	

Preface

This work started somewhere around 2002 when I joined a project, inspired a simple headset for augmented reality audio. The work has evolved around this curious headset ever since, and finally resulting in this thesis work.

First of all, I would like to say my warmest Thank You to my supervisor prof. Matti Karjalainen. I feel truly honored to have had the opportunity to work with him, and grow as a researcher under his supervision and endless resources of knowledge and experience in all areas of acoustics and audio. I would also like to acknowledge my instructors Dr. Aki Härmä, whose way of getting things done was always inspiring, and Docent Ville Pulkki, whose ability to focus on the essential in his work is something that I wish to learn some day. I would also like to thank my co-authors Ville Riikonen and Marko Hiipakka for fruitful co-operation, and also Julia Turku, Sampo Vesa and Heli Nironen for the enjoyable Kamara-years. Furthermore, I'd like show my appreciation to the pre-examiners Juha Merimaa and Riitta Väänänen for valuable comments and suggestions for improving this work.

I'd like to express my appreciation to Lea Söderman for all the effort she has put into making things just work, and also thank you for the delicious dishes we've been privileged to enjoy in our gatherings. I'd also like to thank Jussi Hynninen for all the help with computers, and also for the great photos from our memorable happenings. Henri Penttinen has been responsible for luring me to so (too) many joyful side-projects, and even made me buy a bass guitar, for which I'm evermore grateful. I would like to thank Tuomas Paatero for true friendship and for genuine off-scheduled lunches, and Jyri Pakarinen for insightful coffee break discussions.

Acoustical Society of Finland has been a big part of my whole "acoustics life" and I'd like to acknowledge all the people I've had privilege to work with in the meetings and happenings. Especially, I'd like to thank Docent Tapio Lokki for all the support and enjoyable cooperation during the last eight years, and Juha Backman for being the ultimate audio reference throughout the years.

The famous acoustics lab atmosphere is really something unique, and I'm so proud and happy that I've had the opportunity to be part of it all these years. I'd like to say my grooviest thank you to the mightiest, or at least the most hilarious Big Band I have taken pleasure in playing with: Vesa, Unski,Paavo, Jukkis, Jykke, Mara, Mikkis, Juha, Tontsa, Carlo, Hanna, Tomppa, Cumhur, Seppo, Antti, Heidi-Maria, Jussi P, Hannu, Tuomo, Mairas, Jussi R, Sami, Klaus, Poju, Jouni ,Poju, Marko, Tapani, Olli, and all the others I forgot to mention.

All the years I have worked in the the Laboratory, I have had the privilege to share a room with an ingenious gentleman, Toomas Aaltosaar. I don't have the words to express how much I've enjoyed our conversations, last minute runs to the cafeteria, and all the shared life that has flown during these years. Thank you Toomas.

Along the course, the work has been part of different projects, and I would like to acknowledge Nokia Research Center, the Academy of Finland, and TKK (MIDE) for funding this work

I'd like to show my appreciation for my parents and my family for all the support. And finally, I'd like to show my deepest love and respect for my precious wife Tiina and my two wonderful children, Helmi and Elias. You are my inspiration.

13.10.2009
Miikka Tikander

Contents

List of Publications

This thesis consists of an overview and of the following publications which are referred to in the text by their Roman numerals.

I Marko Hiipakka, Miikka Tikander, and Matti Karjalainen, "Modeling of external ear acoustics for insert headphone usage", *Helsinki University of Technology, Department of Signal Processing and Acoustics, Report 12. Submitted for publication to the Journal of the Audio Engineering Society (JAES).*

II Miikka Tikander, Matti Karjalainen, and Ville Riikonen, "An augmented reality audio headset", In *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx-08)*, Espoo, Finland, September 1-4, 2008.

III Miikka Tikander, "Usability issues in listening to natural sounds with an augmented reality audio headset", *Journal of the Audio Engineering Society (JAES)*, Vol. 57, No. 6, June, 2009.

IV Miikka Tikander, "Sound quality of an augmented reality audio headset", In *Proceedings of The 8th International Conference on Digital Audio Effects (DAFx'05)*, Madrid, Spain, 20-22 September 2005.

V Matti Karjalainen, Miikka Tikander, and Aki Härmä, "Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources", in *Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP'2004)*, Montreal, Canada, 17-21 May, 2004.

VI Miikka Tikander, Aki Härmä, and Matti Karjalainen, "Binaural positioning system for wearable augmented reality audio", in *Proceedings of the IEEE*

Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA'03), New Paltz, New York, USA, October 2003.

Author's contribution

Publication I: "Modeling of external ear acoustics for insert headphone usage"

Author contributed to writing the literature review and the introduction, and also to constructing the measurement devices. Furthermore, the present author took part in the data analysis of the measurement and modeling results.

Publication II: "An augmented reality audio headset"

The early design process for the headset and the mixer, as well as the plan for the usability experiment, was done in collaboration with all of the authors. The mixer, presented in the paper, was further designed and constructed by the third author, Conduction of the usability test, data analysis and writing of the paper was done by the author of this thesis.

Publication III: "Usability issues in listening to natural sounds with an augmented reality audio headset"

The present author designed and conducted the experiment by himself. Data analysis and writing of the manuscript was performed by the present author, as well.

Publication IV: "Sound quality of an augmented reality audio headset"

The author of this thesis designed, implemented and conducted the listening test. The data analysis and the writing of the article was done by the present author, as well.

Publication V: "Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources"

The present author contributed to writing the chapter on position and orientation estimation. All the laboratory measurements were performed together with the first and the present author. Implementation and data analysis was performed by the first author.

Publication VI: "Binaural positioning system for wearable augmented reality audio"

The original idea of the method, introduced in the paper, came from the third author. The present author implemented the algorithm, performed the laboratory tests and analyzed the data. Manuscript was mainly written by the present author with some contribution from other authors.

List of Abbreviations and Symbols

AR Augmented Reality

ARA Augmented Reality Audio

CIC Completely in Canal (hearing aid)

DOF Degrees Of Freedom

FPGA Field Programmable Gate Array

GPS Global Positioning System

HMD Head Mounted Display

HRTF Head Related Transfer Function

IA Interaural Attenuation

MARA Mobile Augmented Reality Audio

RFID Radio Frequency Identificatino

TDOA Time Delay of Arrival

UWB Ultra Wide Band

WARA Wearable Augmented Reality Audio

WLAN Wireless Local Area Network

1 INTRODUCTION

In this day and age, many people carry digital appliances with them continuously. Mobile phones are no longer just telephones but access terminals for all kinds of information channels. One can set up a meeting with multiple people even if each participant is located at some different geographical location. Many of the devices, such as cameras, music players, mobile phones, laptops, etc., have wireless network capabilities and can be connected to a network when in range. This enables various application scenarios for enriching currently existing services. For example, while at the movies an individual can check with their mobile terminal whether the cafeteria next to the movie theater is open or not.

Thanks to the wide use of GPS-enabled mobile devices, location-aware applications are gaining popularity among users. Information can be automatically provided to a user based on his/her location. For example, when entering a cafeteria, a user could automatically receive the menu of the day on their mobile terminal. However, the above scenario still requires that the user maps the information to their local environment. Currently, this kind of information is given merely based on physical location data, and thus the service is not aware of the user's current focus of attention, unless specifically given. If the user's head orientation could be estimated then information could be processed and presented more relevantly.

One large application area exploiting user position and orientation data is augmented reality (AR) where the user's environment is enriched with virtual objects, such as sounds and images [4, 5]. In AR the user perceives the natural surroundings just as they are, but with the use of special hardware, virtual objects are placed in the user's natural environment. To provide visual augmentation, see-through video display glasses are required. Likewise, to enrich the audio environment a special headset is needed. At one extreme, an augmented reality system could be designed so that the user cannot separate virtual objects from the natural surroundings, whereas at the other extreme virtual objects could be intentionally made to stand out. For example, a user in a cafeteria scanning free tables next to the window could be given information about table reservations regarding each table. Depending upon which table the user is looking at, reservation data could be presented as a spoken virtual note, or as a virtual visible note placed on each table.

This thesis work concentrates on the audio-related issues in AR (Augmented Reality Audio, ARA). Furthermore, the emphasis in this work is focused towards mobile usage of augmented reality (Mobile Augmented Reality Audio, MARA), which sets different requirements for hardware and usability of the system when compared to fixed laboratory settings. In literature, different terms have been used to refer to mobile usage of ARA, e.g., *mobile*, *handheld* [61] and *wearable* augmented reality audio (WARA). However, in this work the term *mobile* is used to imply small and practically portable devices. The concept of Augmented Reality Audio is more thoroughly introduced in Section 2.

In ARA, a special headset is needed to provide the user with virtual services. A basic user-worn setup consists of a pair of earphones with integrated microphones for capturing and reproducing sounds, and a unit for computing and handling the communication between a user and the central ARA infrastructure. Of course, any or all of the above components can be embedded in already existing devices, such as mobile phones, or perhaps, the whole system could be integrated within a small wireless earphone. Overall, the concept design in this work is driven by a vision of a device and infrastructure that provides an eyes-free, hands-free, continuous, and ubiquitous experience in ARA.

One prerequisite for ARA applications is a head tracking and positioning system. This is since the ability to track a user's orientation in real-time is needed for aligning the virtual objects with the real surroundings. Currently, many devices exist with integrated GPS systems and are extensively used, for example, in navigational systems. This technology can be exploited for ARA applications as well. However, GPS-based systems do not function well in-doors and, at the time of writing this thesis, no standardized technology exists for indoor positioning. Furthermore, the lack of practical head tracking technology is one of the main obstacles when creating true ARA applications for mobile usage.

ARA technology can be exploited by many different application areas ranging from telepresence to entertainment. However, the majority of the applications will most probably be targeted for common people and for common daily activities. For example, having a meeting with a group of people where virtually present participants have distinct locations in relation to physically present people would be warmly welcomed by many without doubt. One idea behind ARA is that a user would be wearing the ARA system continuously on an everyday basis, and virtual information layered on the real world would be a natural way of retrieving and sharing information.

This raises a question whether the users will be willing to wear the system and perceive their everyday life through an ARA system throughout the day. Naturally, the sound quality of the system is an important factor but also other, non-technical, issues will arise if this kind of technology will emerge. The only way to determine what aspects are relevant in the overall acceptability of ARA systems, and what is to be taken into account when designing the hardware and applications for an ARA, is to test actual ARA systems in real-life situations.

This thesis work attempts to answer questions related to ARA as discussed above and hopefully pave the way for practical hardware and applications. Much work still remains to be done but hopefully a world full of potentially useful ARA applications will make the effort worthwhile.

1.1 Aims of the thesis

The aim of this work is to introduce novel methods for head tracking and positioning and apply them to augmented reality audio. Furthermore, questions related to the design of an audio platform for ARA is studied, and based on the results, a mobile platform for ARA usage is developed and evaluated. Another aim of this work is to evaluate the overall usability of ARA technology, especially in real-life situations.

The results of this work can be used to further develop and refine current ARA hardware, and also recognize the different aspects that must be taken into account when designing ARA hardware and applications applied to real-life situations.

1.2 Organization of this thesis

This thesis is organized as follows. In Section 2 the concept of ARA is defined and some possible application scenarios are reviewed. Section 3 introduces the acoustic problems related to designing headsets for ARA, and in Section 4, the evaluation of sound quality and the usability of an ARA hardware is introduced. In Section 5 the basic scenarios for positioning and head tracking for augmented reality are reviewed, and the concept of acoustic positioning is introduced. The publications included in the thesis work, are summarized in Section 6, and finally, in Section 7 some conclusions of this work are drawn.

2 AUGMENTED REALITY AUDIO

Augmented Reality Audio (ARA) is a concept where a person's natural surroundings is enriched with virtual sounds. The possibility to perceive the natural acoustic environment around a user differentiates it from the traditional concept of a virtual environment where the user is typically immersed into a completely synthetic acoustic environment.

One of the commonly used definitions for AR is given in [4] stating that an AR system should have the three following characteristics:

1. Combines real and virtual
2. Interactive in real time
3. Registered in 3-D

This definition is applicable to ARA systems as well. All of the three characters set different requirements for ARA system and hardware. Especially for mobile usage there are many challenging technical problems to be tackled before a practical (and mobile) system can be built, head tracking being among the most challenging ones.

2.1 Real, virtual, and augmented audio environments

The basic difference between real and virtual sound environments is that virtual sounds are originating from another environment or are artificially created, whereas the real sounds are the natural existing sounds in the user's own environment. Augmented reality audio combines these aspects in a way where real and virtual sound scenes are mixed so that virtual sounds are perceived as an extension to the natural ones.

2.2 Interaction with ARA

ARA applications can be designed so that they do not interfere with the user practicing other activies, i.e., the application leaves the user's hands free and does not require visual attention from the user. However, many applications still require some sort of interaction from the user. Naturally, a mobile terminal could be used for interacting with the application but this would require the user to move the attention from any other activity to the mobile terminal. For an introduction to interfaces for mobile AR see, e.g., [37].

For obtaining the hands- and eyes-free functionality, the interaction should be handled with sound- or gesture-based systems. The interaction can be divided into two categories based on the direction of information flow: a user receiving information from an ARA application, and a user giving information or responding to an application.

In ARA the main interface for giving information to the user is the audio playback system. This kind of an audio-only way of conveying information is called *Auditory display* [9]. There are many ways for giving information to the user through an auditory display. The information can be given as recorded or virtual speech, non-speech sounds, such as earcons or auditory icons[1], or a combination of all of these (see e.g. [60, 13, 16, 15, 75, 44, 50]). Furthermore, in addition to the sound signal design the auditory display can be spreads around the user in 3D [64].

[1]Earcons are structured sequences of synthetic tones that can be used in different combinations to create complex audio messages [16], whereas auditory icons are everyday sounds used to convey information to the user [23]

When the user has to respond or give information to an ARA application, the sound-based interaction can be given via the microphones in the headset. Speech recognition can be used for voice-based controlling of an application. In addition to speech, also other sounds can be used to control ARA applications [45, 77]. When head tracker data is available, head gestures can also be used for interacting with services.

2.3 Rendering schemes for ARA environments

There are different possibilities for producing and mixing the virtual sounds with natural surroundings. The two obvious ways would be either by using a personal headset or by an external loudspeaker setup.

Headphones are a practical solution for mobile ARA because the sound system is carried by the user wherever he or she goes. A headphone-based system also enables providing ARA applications and services without distracting others. A common way to render audio around the user in headphone-based augmentation is to use Head Related Transfer Functions (HRTF) [49]. In this method the sound transmission from a certain point in the environment to a specific point in the external ear, for example the ear canal entrance, is measured. By simulating this transmission by, for example, HRTF filtering, an illusion of externalized and 3D-positioned sound can be reproduced. This technique works well with individualized HRTFs and with carefully placed and equalized headphones [32]. However, finding a set of generic HRTFs that would work with a number of people has turned out to be quite a challenge.

Another challenging problem with headphone based augmentation is that for rendering static sound objects in the environment the user's position and orientation need to be estimated in real time. With real-time head tracking the virtual sound objects can be tied to the surrounding environment and thus the virtual environment stays in place even if the user moves. Another benefit with real-time head tracking is that the dynamic cues from head rotation help localizing sound objects in the environment. Especially for front-back confusion head tracking has been found helpful [11].

Another way to augment the sound environment is to use an external loudspeaker setup. This could be, for example, a room with speakers on the walls. As the loudspeakers are statically placed in the environment, the augmentation is not dependent on the user position or orientation. However, most of the multi-channel loudspeaker reproduction systems suffer from the sweet-spot problem [24, 54, 55], i.e., the accuracy of spatialization worsens when the listener moves further away from the sweet-spot. With location tracking this phenomenon can be compensated for but the drawback is that it only works for this specific user, whereas for the other users in the same space the spatialization accuracy gets worse. Even though there is no need for head tracking for keeping the sound environment static, some positioning system is still needed for location based applications. Another drawback with loudspeaker augmentation is that it is very hard to create personal, i.e., sounds statically around the user, or private augmentation in public places.

Both loudspeaker- and headphone-based ARA system have their advantages and disadvantages. Therefore, in practice, real ARA systems will included both approaches for augmenting the environment, and both methods should be supported by the ARA devices and applications. However, this work concentrates more on the mobile usage of ARA, and therefore the main emphasis is also kept in headphones-based systems.

2.4 Pseudoacoustics

Augmented reality audio can be reproduced to the user with a special headset designed for the purpose. In most cases the headset interferes with listening to natural surroundings by blocking the ear canal to some degree. In an ARA headset a pair of binaural microphones are integrated with the earphones, and by feeding the microphone signals directly to the earphones the user is re-exposed to the natural sounds of the surroundings. This is very similar to using a hearing aid for listening to the environment. In this thesis work, to separate this from natural listening (without headphones), the sound environment heard through the headset is called *pseudoacoustics*.

One of the main targets in designing a headset for ARA usage is to maintain the pseudoacoustics as similar as possible to the natural hearing experience without headphones. More on the headset design is explained in Section 3.

2.5 Application scenarios for ARA

This section reviews some application scenarios related to mobile usage of ARA. For an extensive review on AR applications in general see for example [5, 4].

Current social network and micro-blogging services (e.g. [21, 73]) could be extended by the application scenarios introduced in [2], where Ashbrook and Starner proposed a system where the users' GPS data was studied and a model was created to predict user movements and significant places. This kind of a system would be aware of predicted locations of other users, and with this information the user could be, for example, prepared and notified of upcoming situations.

Rozier *et. al.* [59] proposed an ARA system called *Linked Audio*, where a user can create ”audio imprints” at specific locations outdoors. The imprints can consist of layers of music, sound effects, or recorded voice. When other users reach the specific area, the imprint is played to the user. Different imprints can be linked together, and once a linked imprint is found the user is guided from the imprint to other related imprints in the area. In [31] Härmä *et. al.* proposed an application called *Audio post-it*, where a user can leave acoustic notes in specific locations. When other users come to the reach of the note position, the note is played to the user. A similar application, *ComMotion* was suggested in [46], where a to-do list was played to the user when he or she approaches the task-related area.

One of the most promising ARA applications is binaural telephony [31], which can be extended to a full-blown ARA meeting. Participants of the meeting do not have to be physically present in the meeting but rather they can be present remotely via the communication network. Each participant can place other participants as they wish in their own surroundings. The proposed ARA system setup for the application, and for ARA in general, is shown in Figure 2.1. This is also the system platform that has been used as a basis for the work introduced in this thesis.

There are many task areas, such as driving, surveillance, gaming, etc., where it is important that the user's visual attention is not distracted. In [20], Dicke *et. al.* compared audio-only interface to visual interfaces in driving conditions. In general, both interface types were found to function equally well for operating the system. However, the audio-only interface was found less distracting for driving performance, thus resulting in fewer errors in driving performance. Another eyes-free application, *Auditory Calendar*, is proposed by Walker *et. al.* in [82], where the events in a personal calendar are rendered around the user based on the scheduled times. A meeting at noon would be announced in front of the user, whereas an announcement for an event at 3:00 p.m. would be rendered to the right.

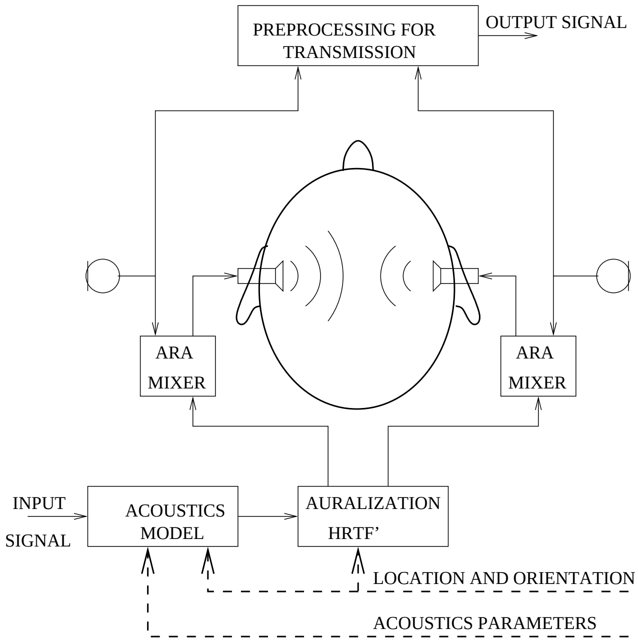

Figure 2.1: An ARA-system introduced in [31].

One application area in ARA are audio-only games [3]. In [57] Röber and Masuch present an overview of audio-only games, and discuss the methods and techniques to play and design such auditory worlds. One example from their review is a classic *Frogger game* where the user has to cross a busy street, and while crossing the street the user has to avoid getting hit by a car.

The microphones in the headset enable analyzing the sound environment around the user. This opens another line of possible applications. The headset of a remote user could be used for surveillance purposes, or locally the ARA system could be used for analyzing the surrounding environment [58, 30], and possibly warn the user of approaching vehicles. And, if someone is interested in identifying birds, for example, the system could be tuned to look for different bird species in the environment [22] and announce the user whenever a new species is detected.

Another important scene analysis area for ARA applications is how to estimate acoustic parameters of surrounding spaces [76, 78]. In mobile usage of ARA, in most cases it must be assumed that the acoustic environment is not known a priori. For convincing augmentation the virtual sounds should be rendered with the acoustic characteristics of the surrounding space, and if the acoustic parameters are not available by other means they must be estimated from the microphone signals.

3 HEADSET ACOUSTICS

An ARA headset has two main functions. The headset should function as an audio interface for ARA applications, and also, it should provide the user with an acoustically transparent reproduction of the natural sound environment. The main difference between the two tasks is how the reproduced sounds should be equalized in order to provide a natural sound reproduction with the headset. For pseudoacoustics, the equalization should compensate for the artifacts created when the headset is placed in the ear as a part of the sound transmission path. The equalization for audio playback, on the other hand, should create an illusion that the sound signal is originated from a certain place in a certain environment.

Designing a headset frequency response is strongly dependent on the headset type. Roughly, the headphones can be categorized by their type of wear and by the type of transducer technology. Figure 3.1 illustrates some examples of different commonly used headphone designs. For ARA usage, most of the headphone types can be applicable. However, for mobile usage the practicability of wearing sets some restrictions on usability of different types of headphones. Also, with ARA the wearing conditions differ from traditional listening of headphones as the usage duration may be fairly long and also the background noise conditions may vary considerably. A good example of a challenging headset environment is airline entertainment headsets as studied by Gilman [25].

The most popular transducer type with headphones has been, and still is, the moving coil transducer [53], which is a miniature version of a traditional dynamic loudspeaker. The design is robust and well understood. The technology enables designing a high fidelity headphone covering the whole frequency band with a reasonable price.

Balanced armature transducers, traditionally more common in hearing aids, have gained popularity in in-ear (insert) headphones. The small size of the transducer is favorable in small in-ear designs. The size of the membrane and the maximum displacement are fairly small, and thus the design quite often suffers from weak low frequency reproduction.

More recently, different kinds of headphones based on bone-conduction have appeared in the consumer market [38, 41, 12, 67]. As the sound is conducted to the hearing system via a mechanical bone conduction, for example by placing the ear peace against the mastoid bone, the technology allows designing headphones that leave the ear canal opening free. This enables an undistracted and natural listening setup for the surrounding sounds.

The main difference between air-conducted and bone-conducted sound is that the sound is transmitted to the basilar membrane through different physical and mechanical sound paths. With traditional (air-conduction) headphones the sound is fed to the hearing system through the ear canal, whereas, with bone-conduction headphones the sound is fed through bone and tissue conduction. In either case the characteristics of the sound transmission paths from the ear piece to the actual hearing perception should be compensated for when designing the headset frequency response. The same equalization targets cannot be automatically applied for bone conduction headphones as used with conventional headphones [39, 83, 66].

One of the main drawbacks with bone-conduction headsets is the very poor interaural attenuation (IA) due to mechanical transmission of sound waves [74]. However, there are many studies suggesting that a stereo image could be created with bone-conduction headsets [84, 43, 65].

From the technology point of view, hearing aids and ARA headsets are very similar. However, there is a slight difference in design philosophy. Whereas ARA headset aims at acoustic transparency, with hearing aids the required gain (or amplification) levels are so high that the seek for maximum gain is the primary design criterion, and often the price paid for the gain is a narrowed bandwidth. The bandwith rarely reaches the highest frequencies, but concentrates more on mid frequencies to make speech as clear as possible. Of course, if the ARA headset is to be used as a hearing aid device then the same problems would arise as with hearing aids. There is a big industry and research going on related to hearing aids and most of this technology can be directly used with ARA headsets. Especially the hardware integrated with the modern CIC (completely-in-canal) hearing aids could be used with ARA headsets as well.

In the following the basic equalization schemes are described for audio playback and for reproducing pseudoacoustics. The derivation is based on an assumption of using an insert-type of headset (the rightmost headset type in Fig. 3.1.). For other headset types the equalization should be applied accordingly.

Figure 3.1: Different types of headphones. From left: Circumaural, Supra-concha, intra-concha, and insert type of headphones.

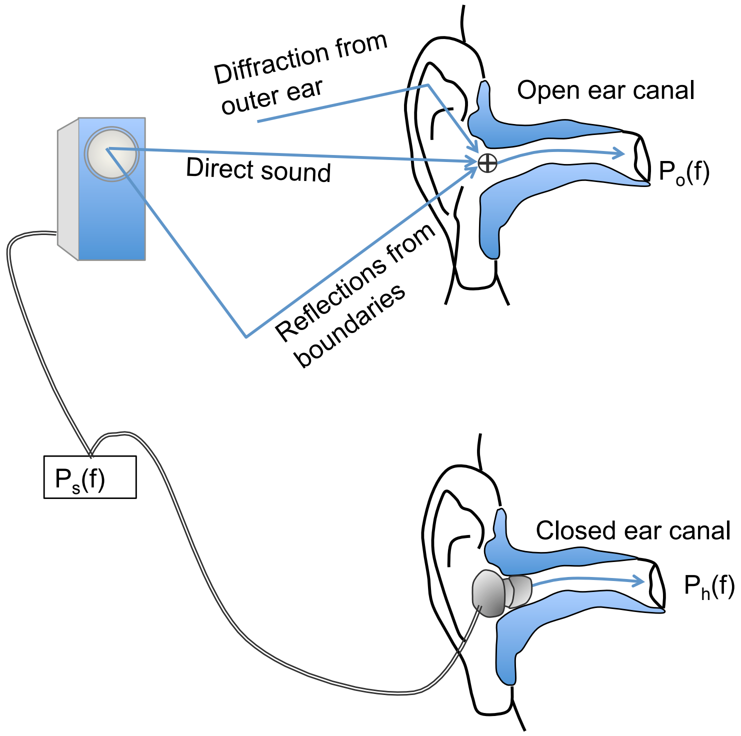

Figure 3.2: A simplified illustration of sound transmission from a source to the ear drum. $P_s(f)$ is the source signal, $P_o(f)$ is the signal at the eardrum in an open ear canal case (above), and $P_h(f)$ is the signal at the eardrum when listening with headphones (below).

3.1 Equalization for audio playback

The sound signal at the ear drum is different from the sound signal that was originally emitted from the sound source. The outer ear (the parts from the tympanic membrane outwards) and the surrounding space modify the sound signals while transmitting sound waves from the source to the ear. Fig. 3.2 shows a simplified comparison of differences in sound transmission paths when listening to sounds in the environment without headphones, and with headphones. In order to reproduce a natural sound perception the headphones should replicate the open-ear sound pressure signals at the ear drum, i.e., $P_o(f)$ should equal to $P_h(f)$.

As an acoustic system, the ear canal can be considered as a quarter-wave length resonator. The resonance characteristics are determined by the geometry of the ear canal and the impedance of the ear drum. For a typical adult ear the first ear canal resonance occurs at around 2-4 kHz [29]. When an earphone blocks the ear canal, this resonance disappears and, as such, the sound field is perceived unnatural. Furthermore, as the ear canal is closed from both ends, it starts to act more like a half-wavelength resonator. This shifts the lowest resonance up to around 5-10 kHz further distorting the sound signal.

Outside the ear canal, the human body (the shoulders, the head, and the pinnae) causes reflections to the impinging sound signal. The modification to the signal due to diffraction from the user's body is called the Head Related Transfer Function (HRTF) [49], which is strongly user-specific and also depends on the angle of sound arrival.

One practical question is, what kind of listening condition should the headphone listening replicate. In addition to the user itself, the reflections from surrounding surfaces add an acoustic signature to the sound signal. Of course, depending on the surrounding space the amount and nature of reverberation vary. In a free-field (anechoic) condition there are no reflections at all, whereas in a big hall most of the sound energy comes from the reverberation field. Commonly there have been two specific equalization targets for headphones, *diffuse field* and *free-field* equalization [48, 80, 69]. Free-field equalization tries to replicate the ear signal in a case where a user is listening to a frontal sound source in anechoic conditions. This equalization scheme is often preferred in studio monitoring situations as it resembles a near-field monitor listening conditions. The diffuse-field equalization, on the other hand, assumes that the sound field is completely diffuse. The most natural environment lies somewhere between these two extreme cases.

When an earphone is placed on the ear, all of the above diffraction phenomena

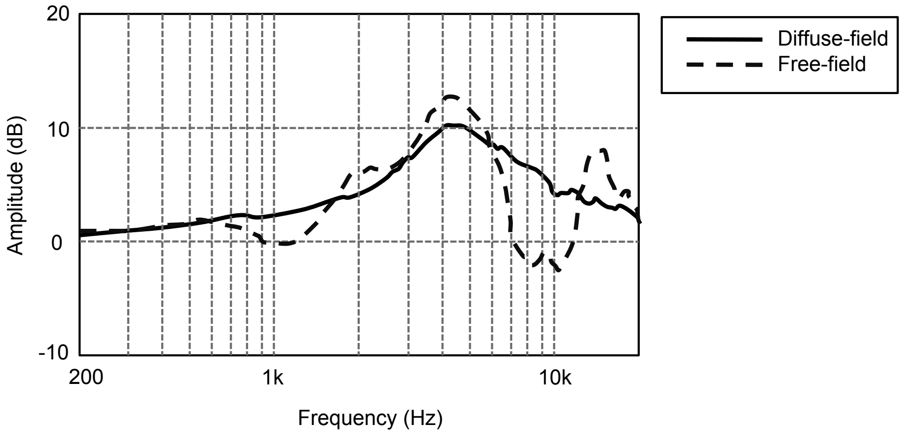

Figure 3.3: Free- and diffuse-field equalization targets for headphones. Measurements are derived from blocked ear canal measurements. Data adapted from [48].

are bypassed, and for natural reproduction, this must be corrected somehow. In the literature there are many studies suggesting optimal equalization curves for different listening situations [48, 47, 40, 10]. Fig. 3.3 shows design targets for free- and diffuse-field equalizations [48]. The responses in the figure are derived from blocked ear canal measurements with a frontal sound source in free-field (anechoic) and in diffuse-field (sounds arrive from all directions) conditions.

3.2 Equalization for pseudoacoustics

Equalization for pseudoacoustics differs from playback equalization. Fig. 3.4 shows the sound transmission path when a user is listening to pseudoacoustics. Comparing to natural listening without headphones, as shown in Fig. 3.2, the sound transmission seems very similar. The spatial cues and diffraction from external parts of the ear are always present in the sound signal, thus for natural sound reproduction these should not be artificially added anymore in the headset or mixer. Only the distortion due to the headset being on the transmission path should be compensated for.

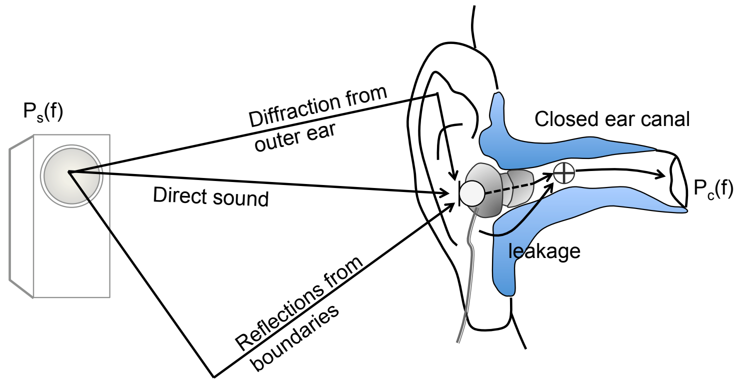

Figure 3.4: Sound transmission from a source to the ear canal with an ARA headset.

The natural sound environment is captured by the microphones outside the earphone casing. In order to maintain the transparent sound transmission, the headset should be equalized so that the sound signals at the ear drum are equal to the ear signals in a natural listening situation ($P_o(f)$ in Fig. 3.2). As stated in the previous section, an open ear canal acts as a quarter-wave length resonator and boosts the sounds in the 2-4 kHz range. When the ear canal is closed by the earphone, the quarter-wave length resonance is no longer existent, but rather, a half-wave resonance is created at around 8-10 kHz [33].

A straightforward way to compensate for the misplaced resonances is to use a resonator filter to recreate the quarter-wave length resonance, and a band-stop filter to remove the half-wave length resonance [56]. This, of course, assumes that the frequency response of the headset is otherwise flat and does not need to be corrected.

Another phenomenon distorting the pseudoacoustics is the sound leakage around and through the headset. Insert earphones, in general, attenuate surrounding sounds quite efficiently in the mid and high frequencies. However, there is always some leakage through and around the headset and also a portion of sound field is transmitted to the ear canal as bone and tissue conduction [34, 35, 71, 53]. Low frequencies can leak through the earphone quite effectively. The leaking from the real environment sums up in the ear canal with the pseudoacoustic representation delivered by the headset. This summing causes coloration especially at low frequencies and deteriorates the pseudoacoustic experience [31]. The amplification of low frequencies has to be equalized, as well as other spectral colorations caused by the headset design itself.

One thing to be noted is that the leakage path has no delay, other than the delay caused by acoustic transmission[2], and therefore only very little latency is allowed also in the electric signal path between the microphone and the earphone. Any latency would result in comb-like filtering distortion in the pseudoacoustic sound environment. In practice this requires using analog circuitry for processing the microphone signals prior to feeding then to the earphones. As for a digital alternative, FPGA -based audio processors might offer an alternative for signal processing as they are capable of very low latency signal processing.

3.3 Occlusion effect

One common complaint with hearing aid users is the boomy or hollow sound of the user's own voice. This is due to the occlusion effect caused by the earpiece closing the ear canal [53]. At low frequencies the acoustic impedance of an open ear canal is mainly determined by the radiation impedance of the ear canal entrance. The low impedance of the ear canal entrance practically "short circuits" any low-frequency sound that is mechanically transmitted to the ear, thus no extra pressure is generated in the ear canal. However, when the ear canal is closed by an earphone,

[2]The distance from a microphone to the transducer in an insert-type of ARA headset is in the range of 1-2 cm. pseudoacoustics is transmitted as an electric signal whereas acoustic leakage travels around the headset with a speed of sound. For 1.5 cm the acoustic travel time is about 45 μs.

which equals to a fairly high impedance [81], the ear canal now resembles a high impedance (pressure chamber). For this reason, any sound that is mechanically transmitted to the ear canal will generate sound pressure that will be heard by the subject. The phenomenon is more pronounced at low frequencies.

This same problem exist with in-ear type ARA headsets as well. One way to overcome the occlusion effect is to use some venting system with the headset, which would lower the acoustic impedance of the headset or simply use more open type headsets. Another way to decrease the occlusion effect, as used with hearing aids, is to use deeply inserted earmolds. When the earpiece reaches almost the ear drum, less sound is transmitted to the ear canal via bone conduction. Though, this solution might be little too unpractical for ARA usage.

3.4 Microphone placement for binaural audio capture

A typical ARA headset configuration has microphones outside of the earphones, and these signals are combined with the virtual sound objects. However, for users to feel comfortable being immersed in ARA space, the perception of the natural sound environment should not be altered by the system.

Depending on the headset type the outer ear geometry is changed from an open ear when the headset is placed in the ear. The microphone placement in the headset should be designed so that the microphones captures as much of the spatial cues as possible. According to Algazi *et. al.* [1, 49], HRTFs measured slightly outside a blocked ear canal still include all the spatial information that would exist at the ear drum in an unoccluded case. Thus, small earplug- or insert-type headsets would still provide all the necessary spatial information. D'Angelo et al. studied how CIC (completely in the ear canal) hearing aids affect the localization ability [18]. A group of normally hearing people were wearing a CIC hearing aid that was equalized to have an identical transfer response compared to an unoccluded ear. In the test the testee's head was fixed and there was no visual clues available. According to results there was a small degradation in localization ability. Brungart *et. al.* [17] performed a similar study with different types of insert hear-through headphones. It was found that the wide enough bandwidth of the system had more effect on the localization accuracy, than microphone placement. The subjects were allowed to move their head during the test. With ARA headsets, the users are also able to move their head while listening to the sound environment, thus being exposed to real-time acoustical and visual cues together with pseudoacoustic signals.

The results presented in Publication III also reveal that a good spatial accuracy can be obtained even with non-individual equalization of headphones. In the study, subjects wore an ARA headset, where a generic equalization was applied to provide a natural reproduction of surrounding sounds. The spatial accuracy was graded very good by all the subjects. Humans adapt to modifications in hearing system fairly rapidly [42, 62, 63], which is actually quite essential as our hearing environment is changing all the time for example by using hats or standing close to a surface.

3.5 Contribution of this work to ARA headset acoustics

In Publication IV, first studies on optimal equalization for an insert-type ARA headset were performed. A generic equalization target curve for insert-type headsets was subjectively introduced. In Publication II the work was further continued and based on the results, a generic system for ARA applications comprehending a headset and a mixer unit, was introduced. Also, the results of a preliminary test on overall quality (sound and usability) of the system in real-life situation was presented. A more thorough test on the sound quality of the system was performed and documented in Publication III. In the study, a group of test subjects evaluated the sound quality of the ARA system in laboratory conditions, and also in real-life situations.

Furthermore, the adaptation to the pseudoacoustics in lengthened use was studied. The acoustic behavior of the external ear together with an insert type headphone was studied in detail in Publication I. The effect of different parameters, such as the length of the ear canal, acoustic impedance of the headset and the impedance of the ear drum, was studied. Based on the results, a physical model of the outer ear and the headset was also introduced.

4 USABILITY OF ARA HEADSETS

Most of the application scenarios mentioned in this work and in the literature are aiming at usage in everyday life situations. Applications, such as virtual tourist guide or getting augmented product information in a grocery store, mean that users are expected to wear an ARA system in the practical activities of their normal life. The usage is mostly voluntary and motivated by the benefits provided by the applications and services. Therefore it is essential that the usability of the system is in balance with the added value gained by the user. This is also what separates the usability question of ARA systems, for example from hearing aids or other health-related devices. With hearing aids the added value, i.e., the ability to hear, overpowers minor usability issues and thus people are very motivated to wear them [26, 51], despite of possible usability issues, whereas with ARA applications users are free to choose whether to wear the system or not.

As a design point of view, designing an ARA headset for high-quality sound is a good starting point. Most probably the users want to hear the surrounding sounds as unaltered as possible. However, in public and in lengthened use, and especially among the users who are not used to wearing earphones, other issues may rise to be important factors in overall acceptability of the system as well. Grinter and Woodruff studied the preference for a headset type for museum tours [27]. In the study, a group of naive subjects, between 40-60 years of age, were instructed to use different types of headsets for a tour-like situation. It was found that, apart from the sound quality, issues such as ease of use and conspicuousness turned out to be important factors in choosing the headset type. Another, non-audio related factor in usability was pointed out in [52], where Park *et. al.* studied how personal preference affected the usability of a system. It was found that users trust or affection for a brand affects favorably the results of a usability study, i.e., the test subjects had less audio quality related demands for a preferred product.

Currently existing ARA applications are still only a promising possibility waiting for practical hardware and applications, and therefore there is no experience on how the general public will receive the technology, and what will be the probable "killer applications" in the future. The only way to get insight into the matter is to perform usability tests in real-life situations with working prototypes. The outcomes of real-life tests will reveal information on the usability aspects of ARA systems, as well as other non-technical issues that will arise when using the system publicly.

4.1 Contribution of this work to ARA headset usability evaluation

The Publications II-IV study different aspects of usability of an ARA headset. In Publication IV, a perceptually optimal equalization curve for an insert type of headset was measured. Furthermore, the overall sound quality and user acceptance of pseudoacoustics, with an equalized headset, was evaluated in laboratory settings.

The results of a pilot usability study are reported in Publication II. In the study, a group of subjects wore an ARA system in their daily routines and reported their observations in a diary. After the field test part the subjects were interviewed and the diary notes were discussed. Based on the results, the sound quality of the prototype system was found good enough for practical situations, and the main critique arouse from handling noise, and limitations caused by the wires. Also, some other non-technical issues, such as social ambivalence while wearing the headset, were discussed.

Publication III reports the results of another usability study where the usability of the system was further evaluated with a test group. Similar to the test, introduced in Publication II, the subjects wore an ARA headset in real-life situations and reported observations in a diary. Furthermore, the sound quality (timbre, spatial impression, and location accuracy), and the annoyance factors of some practical activities were evaluated at the beginning of the test, and also after the field test period. The evaluations were performed at both ends of the test period to study if any adaptation would occur leading to improvements in perceived sound quality or usability during the field test period.

5 HEAD TRACKING AND POSITIONING

Locating a user's head and tracking its orientation is one of the key points and also one of the major challenges in augmented reality audio. When the users' position data are available, virtual sound objects can be designed to have distinct places in the environment. The audio objects can be rendered in relation to different coordinate systems. For example, some applications may use the natural environment as a base coordinate system for placing the sound objects, whereas some applications could use a user-centric coordinate system as the basis for rendering the sound objects around the user.

Depending on the application there are different requirements for user positioning. Personal virtual objects are produced with the user's own headphones, whereas public augmentation could be created with an external loudspeaker system. If the augmentation is produced with an external loudspeaker system, a plain location data (without head tracking) is sufficient to produce static augmentation. However, personal augmentation (with headphones) requires that also the user's head orientation is estimated in order to enable static augmentation of the environment, i.e., the virtual sound objects stay in place in the environment even when the user moves. Furthermore, there are application areas where only the orientation, without location tracking, might be sufficient for augmentation. For example, when using a vending machine or an on-line banking terminal, the user position could be predicted fairly well, and only the orientation needs to be tracked separately for the augmentation.

The term positioning is sometimes understood as only estimating a subject's location, perhaps due to the popularity of GPS navigation devices. However, the term positioning, in it's full extent, includes locating an object and also estimating it's orientation. This is called 6-DOF (degrees of freedom) positioning. The six degrees of

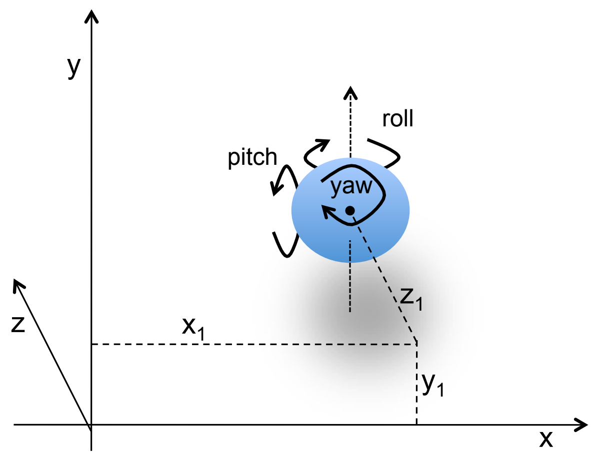

Figure 5.1: Directions of rotational vectors in positioning, when heading, shown by the dotted arrow, is in positive y-direction. z-axis is pointing upwards from the surface.

freedom are the location coordinates, for example, *x, y, z*, and the orientation data, *pitch, yaw*, and *roll*, in three different rotational directions, as shown in Fig. 5.1. For ARA the object to be positioned is the user's head.

Nowadays, estimating the location, especially outdoors, is fairly straightforward. Practical and affordable GPS units are already available, and in increasing numbers embedded in existing mobile devices. With GPS the positioning is performed with a help of the satellite system in the orbit of the earth. However, the general GPS-based systems do not work well indoors, and other means are required for indoor use. In addition to GPS there are also other similar systems for global positioning. Table 5.1 lists the currently existing most common global positioning systems.

For indoor positioning there are still no practical solutions available, as far as mobile ARA usage is considered. Different kinds of custom and ad-hoc systems have been developed, but none of the methods offer the usability, for example, compared to current GPS systems. Most common technologies for indoor (or local) positioning

Techology	Maintained by
GPS	USA
Galileo	European Union
Glonass	Russia
CNSS (BeiDou 1)	China
WAAS	USA, based on GPS

Table 5.1: Most common global positioning systems utilizing a satellite network.

are listed in Table 5.2.

Compared to location tracking, head tracking is the harder part in the positioning problem, especially for mobile AR. Position tracking benefits from the fact that the positioning device can be carried quite freely, for example in a pocket or in a bag, and still get good positioning accuracy. Head tracking on the other hand, requires the device to track the user's head position and orientation in order to be usable in augmented reality audio. In practice this means that the device should be attached somewhere around the user's head. In ARA, one possible place for attaching the head tracker is the headset that is carried by the user anyway. The obvious benefit in integrating the head tracker with the headset is that the headset automatically follows the user's head orientation. However, the small size of the earphone might pose some challenges for designing a head tracker in a sufficiently small casing.

For augmenting the natural surroundings, the head tracker should be able to offer absolute coordinates of the user continuously, and in real-time, for keeping the virtual sound objects in their right places in the environment. Furthermore, it is essential that the positioning and head tracking data does not drift over time [36].

Most of the head tracking research has been motivated by visual augmentation of

Techology	Available position data
Video based	location, orientation
Acoustic	lcoation, orientation
Extented GPS	position
WLAN	position
Bluetooth	proximity
Wibree	proximity
RFID	proximity
UWB	proximity
Zigbee	proximity
Magnetic	location, orientation
Accelometers	location, orientation
Gyros	location, orientation

Table 5.2: Technologies for indoor positioning.

surroundings, and the AR is often provided with a see-through HMD (head mounted display) [5, 85]. The required accuracy for aligning virtual visual objects with real visual surroundings is much higher compared to aligning virtual audio objects with the natural surroundings, especially when sound is used to augment real physical objects. Even the smallest misalignments in visual augmentation can be easily noticed. For audio, the visual cue of a real physical object helps considerably to make virtual sounds perceived to originate from the target object [11].

There is no single technology that would give an answer to an all-around positioning and orientation tracking system. Therefore, a practical system will take advantage of multiple sensors (sensor fusion) and networks to estimate the user's location and orientation, based on currently available data (data fusion). This is called hybrid tracking [8, 6, 28, 85, 68, 86]. In addition to sensor data, the positioning system could also try to predict the user movement, based on human physical constraints and movement models [7] or by a user navigation model [2].

5.1 Acoustic positioning and head tracking

Acoustic positioning and head tracking is one way of estimating the user position, among the methods mentioned above. Using sound signals for detecting objects is nothing new, and the method has been widely utilized, for example, in underwater radars. An overview of different acoustic positioning systems is given in [79].

For ARA usage, there are mainly two basic approaches for acoustic positioning. Microphone arrays can be placed in the space in known locations, and by analyzing the captured sound field the person within the area can be tracked. Alternatively, known sound sources (anchors[3]) can be placed in a room and the sound sources can be detected with the microphones within the ARA headsets [72]. For practical usage the anchor signals must be designed so that they are not (disturbingly) audible to the users.

Using a microphone array to detect the user movement requires the user to wear a sound source for the tracker system to "hear" the user continuously. The tracking is performed based on detecting the microphone signals in the array, and as a result the system gives the most probable direction for the sound source in relation to the array. The benefit of this approach is that the user-worn system only needs to emit the tracking signal, and the infrastructure takes care of the signal processing. For separating different users, the signals should be individualized somehow. There is a lot of literature on signal processing techniques for microphone arrays, and for an overview, see, e.g. [14, 19].

[3] In literature, the term beacon is often used for noting signal-emitting devices in the environment, used for navigation purposes.

Another way to perform acoustic tracking is to turn the microphone array setup the other way around and let the user wear the array, which is used to track sound sources in the environment. In ARA, the user is wearing a headset with an integrated microphone pair (a binaural microphone pair/array) that is continuously following the user's location and orientation. By placing static sound sources, known as anchor sources, in the environment the binaural microphone array could be used to detect the sources, and to estimate the user's location and orientation [58, 72]. The signals for the anchor sources could be specifically designed for the tracking system, or alternatively some other unknown signals with a distinct spectral, for example ventilation channel noise, could be used as well.

Acoustic tracking has its drawbacks, as well. Naturally, when using anchor signals within the audible frequency range, the signal levels must be kept low. This, of course, decreases the signal-to-noise ratio of the whole system, which lowers the accuracy of the tracking system. One idea to improve the situation is to use smart anchors that, for example, filter the anchor signal so that it is masked by the background noise. Using anchor signals that are above or at the limits of a human hearing threshold enables increasing the anchor signal levels. However, high frequencies have the downside that they are easily shadowed by even the smallest objects. Especially the user's head may shadow the microphone that is further away from the sound source. One way to overcome this is to have enough anchor loudspeakers (or microphone arrays) in the environment, so that several of them is always visible to each ear.

Figure 5.2 shows a schematic view of a generic positioning setup with *n+1* loudspeakers in the environment and a user wearing a binaural microphone pair. For basic positioning, for example, starting with the anchor *0* at position (x_0, y_0), an anchor signal is played from the anchor loudspeaker and then captured by both microphones. If the playback and the recording systems are in synchrony the positioning system enables estimating the signal propagation times from the anchor to both microphones ($t_{0,l}$ and $t_{0,r}$ in Fig. 5.2), which can be further converted to distances. If there is no synchrony between the playback and the recording system, only the angle of arrival can be detected by calculating the time delay of arriving (TDOA) signals between the microphones ($t_{\text{diff},0}$ in Fig. 5.2). When this is repeated for multiple known anchor sources, the user position and orientation can be estimated, for example, by trilateration [70].

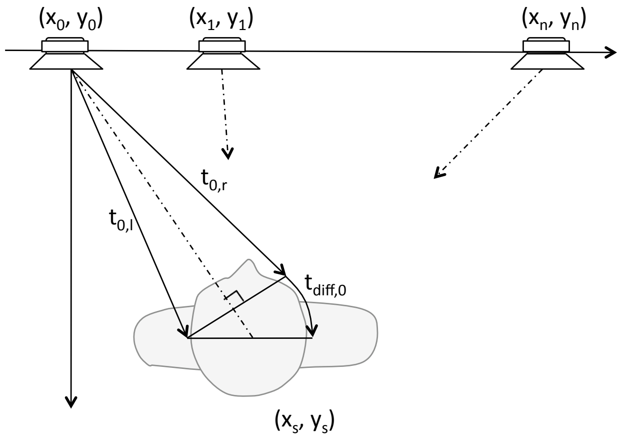

Figure 5.2: Schematic view of an acoustic positioning system.

5.2 Contribution of this work to head tracking and positioning

Publication VI introduces a novel method for head tracking and positioning for AR usage. The method utilizes a pair of binaural microphones worn by the user (integrated with the ARA headset), and a number of anchor sound sources in the environment. The sound sources emit a known sound signal that is captured by the headset microphones. By estimating the time delay (via cross-correlation computation) between the emitted and captured sound signals, the orientation in reference to the loudspeaker can be estimated. Different loudspeakers (anchors) are separated by dividing the anchor signal in smaller frequency bands, or by polling the loudspeakers with the anchor signal.

In Publication V the acoustic positioning system was further developed by introducing new methods for sound signaling between the anchor sound sources and the microphones. The usage of high audio frequencies for anchor sources was introduced. Using high (> 10 kHz) frequencies has many advances, such as background noise being less disturbing, the user's own speech does not interfere with the signals, and also the highest frequencies are inaudible to humans. One idea, also introduced in the publication, is a method, where the anchor signal is modulated to a higher frequency range. With multiple anchor sources each loudspeaker can be separated using different modulation frequencies for different loudspeakers. Signal from each loudspeaker is revealed by demodulation by multiplying the microphone signals with complex-valued carrier followed by low-pass filtering.

6 SUMMARY OF PUBLICATIONS

This thesis comprises six publications, two of which were published in international reviewed journals and four in international full-paper reviewed conferences. The articles are divided into two main topics. Publications I - IV discuss sound quality and usability as well as modeling of an ARA headset, and publications V - VI address binaural acoustic positioning for AR applications.

Publication I: ”Modeling of external ear acoustics for insert headphone usage”

This papers studies the effect that the ear canal length and the ear drum impedance have on the earphone-generated sound pressure inside the ear canal. Two different ear canal simulators with adjustable ear canal lengths were constructed, one with a hard and one with a damped ear drum simulator. Furthermore, a dummy head with different sized pinnae, to be used together with the ear canal simulators, was constructed to study the effect that the external parts of the ear have on the sound pressure inside the ear canal. The measurements were also repeated with real ears to validate the simulator measurements. Based on the measurements a physical model of the earphone to ear canal interaction was formulated.

Publication II: ”An augmented reality audio headset”

This paper introduces a generic ARA system, comprehending of a headset and a mixer, and also a preliminary usability evaluation of the system. The mixer unit enables individual equalization for different users. For minimum latency, the system was built by using analog circuitry. The usability of the prototype system was evaluated with an evaluation performed in real-life situations. A small group of subjects wore the headset from four days to a week in everyday life situations. The subjects kept diary on their usage of the headset, and after the field test they were interviewed for comment on the usage and issues that had possibly risen during the field test period. The sound quality of the prototype headset was found good and applicable for practical usage in everyday-life situations. Some usability issues in real-life case are also discussed in the paper.

Publication III: "Usability issues in listening to natural sounds with an augmented reality audio headset"

This paper reports the results of a usability study, which evaluates the overall usability and sound quality of an ARA system (introduced in Publication II). The evaluation was performed both in laboratory settings, and in real-life situations. The test consisted of three parts: sound quality and usability evaluation in the beginning of the test, the field test part in a real-life environment, and the sound quality and usability evaluation at the end of the usability test. The first and the last parts were performed in laboratory settings and the main focus was to evaluate the sound quality (spatial impression, timbre, and location accuracy), and the annoyance factors on performing some practical daily activities, and also to see if any adaptation would happen during the field test period. In the field test, the subjects were given the headset with them and they were instructed to wear the headset as much as possible in their daily activities, and to write down any observations in a diary. This part lasted from four to seven days and during this time each subject wore the headset from 20 to 40 hours.

The sound quality of the headset was found good, and especially the spatial sound qualities were found very good. It was also found that some adaptation to the sound quality of pseudoacoustics occured during test period, whereas for annoyance of practical activities, such as eating, drinking and speaking, no adaptation was discovered. Furthermore, other practical usability issues are discussed in the paper.

Publication IV: "Sound quality of an augmented reality audio headset"

In this paper a subjective measurement was performed to yield a perceptually optimal equalization curve for an insert-type of headset. In the test the subjects listened to sinusoids with a pair of circumaural headphones, and under the headphones, the other ear was left open and the other ear had an insert earphone placed in the ear with a directly coupled miniature microphone (functioning just like a hearing aid). Frequency by frequency the subjects had to match the sound pressure levels on each ear. A group of nine testees performed the measurement, and based on the resulting individual equalization curves a generic target equalization filter was formed. The filter was formed by taking the average of the individual measurement data. The sound quality of the equalized headset was further evaluated for sound quality by the same test group.

Publication V: "Head-tracking and subject positioning using binaural headset microphones and common modulation anchor sources"

The acoustic positioning system introduced in Publication VI was further developed by introducing new methods for sound signaling between the anchor sound sources and the microphones. The usage of high audio frequencies for anchor sources was introduced. Using high (> 10 kHz) frequencies has many advantages, such as background noise being less disturbing, the user's own speech not interfering with the signals, and the highest frequencies being inaudible to humans.

Also, a method to use multiple anchor sources with the same anchor signal simultaneously was introduced. The method was based on modulating the anchor signal using a unique modulation frequency for each loudspeaker. At the reception, signal from each loudspeaker was revealed by demodulation by multiplying the microphone signals with a complex-valued carrier followed by low-pass filtering.

Publication VI: "Binaural positioning system for wearable augmented reality audio"

An acoustic positioning system for ARA was introduced. The positioning system was comprised of a set of known anchor sound sources in the environment and a binaural headset with microphones integrated with earpieces. The time difference of the arriving sound between the microphones was measured by taking the cross-correlation between the known anchor signal and the captured microphone signals. The distance between the peaks in the cross-correlation functions between the microphone channels was used to determine the orientation of the user. The movement of the user was estimated by detecting the movement of the peaks in the cross-correlation functions. For multiple simultaneous anchor sources the sources were separated in the frequency domain by having a specific frequency band for each anchor.

7 CONCLUSIONS

The concept of augmented reality audio was introduced already decades ago and the demos and application scenarios presented in the literature (and in movies, of course) are fascinating. The idea of layering an interactive virtual world on the real world appeals to the imagination of curious minds, and most probably there will be a demand for this technology. The technology for providing and building full-blown ARA systems is already available. However, real applications and systems for common people are still waiting to be discovered.

This thesis work has studied different aspects of ARA, and also introduced methods and techniques for ARA systems that should shorten the path towards practical ARA applications and hardware.

Undoubtedly, one of the main obstacles to widespread use of ARA applications, especially in the mobile realm, is the lack of practical head tracking systems. User position and orientation tracking is essential for many ARA applications, and therefore this is the area that needs to be tackled first. The widespread use of GPS is a good example of what might happen with ARA once base technological issues are solved. At the moment there is a myriad of applications utilizing GPS technology, and the same could be possible with ARA. The binaural head tracking system introduced in this work offers one possible solution to the user head tracking problem. Specifically, in local and smaller area installations, this approach offers a plausible solution.

For a common user, the most tangible part of the ARA system is the headset. Preferably, the headset should be the only ARA device for the user to wear. The headset has a central role in the entire ARA system, as its function is to provide all the audio services for the user, namely producing and capturing sound, and to replicate the surrounding sound environment, i.e., pseudoacoustics. Humans develop continuously in retrieving and perceiving information from their aural environment. Humans can detect where sound events originate from, while certain sounds automatically raise our attention level (for example a baby's cry). This happens automatically within a human's life-long trained hearing system. If an ARA system is assumed to be worn for longer periods in everyday-life, it is obvious that the sound quality of pseudoacoustics must be sufficiently good. Determining what is sufficient is not an easy question to answer since not only does it depend on the user and the equipment, but also on the added value the user gains from an application.

The sound quality of the generic ARA system introduced in this work was found to be good for most practical situations. However, for critical listening situations, such as sound quality evaluation, there is still room for improvement. Every user has a unique pair of ears, and this should be taken into account while tuning the headset system for some particular user. How this should happen is still an open question. Insert earphone acoustics is still a fairly new topic in the area of audio research, though similar topics have been studied in context with hearing aid equipment and earplugs. This is an area that still requires further research. Naturally, the design ideas utilized for ARA headsets could be exploited and applied to hearing aid and headphone design, as well.

Once ARA technology finds its way into widespread use, there will be many interesting issues to be considered. Apart from professional usage, such as medical or military applications, most of ARA usage will probably occur in everyday life situations. ARA has potential to radically change the way in which people communicate and are present with each other. A comparable example is the mobile phone and how it altered the manner of communication and reachability. There will also be many usage etiquette issues to be solved once ARA technology emerges. For example, how to notify others that the user is listening to them (through the headset) and not something else from the headset? The usability studies reported in this work gives some insight into the subject, but more studies are required with working ARA applications in real situations.

ARA, in one form or another, will be an inevitable part of humanity's future way of living. When this will occur still remains to be seen. Meanwhile, we are left to enjoy reality.

8 ACKNOWLEDGEMENT

This work has been supported by the Academy of Finland (project nos. 53537 and 106956), Helsinki University of Technology (project MIDE/UI-ART) and the Nokia Recearch Center (KAMARA projects).

References

[1] V. Ralph Algazi, Carlos Avendano, and Dennis Thompson. 1999. Dependenceof Subject and Measurement Position in Binaural Signal Acquisition. Journal of the Audio Engineering Society 47, no. 11, pages 937–947.

[2] Daniel Ashbrook and Thad Starner. 2003. Using GPS to Learn Significant Locations and Predict Movement Across Using GPS to Learn Significant Locations and Predict Movement Across Multiple Users. Personal and Ubiquitous Computing 7, no. 1.

[3] Audiogames.net. Accessed: May 13, 2009. http://www.audiogames.net/.

[4] Ronald Azuma. 1997. A survey of augmented reality. Presence: Teleoperators and Virtual Environments 6, no. 4, pages 355–385.

[5] Ronald Azuma, Yohan Baillot, Reinhold Behringer, Steven Feiner, Simon Julier, and Blair Macintyre. 2001. Recent Advances in Augmented Reality. IEEE Computer Graphics and Applications 21, no. 6, pages 34–47.

[6] Ronald Azuma, Bruce Hoff, Howard Neely III, and Ron Sarfaty. 1999. A Motion-Stabilized Outdoor Augmented Reality System. In: Proceedings of IEEE Virtual Reality ’99, pages 252–259. Houston, Texas.

[7] Ronald T. Azuma. 1995. Predictive Tracking for Augmented Reality. Ph.D. thesis, University of North Carolina, Department of Computer Science, Sitterson Hall, Chapel Hill, NC 27599-3175.

[8] Ronald T. Azuma, Bruce Hoff, Howard E. Neely III, and Ronald Sarfaty. 1998. Making Augmented Reality Work Outdoors Requires Hybrid Tracking. In: Proceedings of the First International Workshop on Augmented Reality. San Fransisco, CA.

[9] Stephen Barrass and Gregory Kramer. 1999. Using sonification. Multimedia Systems 7, no. 1, pages 23–31.

[10] B. B. Bauer. 1965. Improving headphone listening comfort. Journal of the Audio Engineering Society 13, no. 4, pages 300–302.

[11] D. R. Begault, E. M. Wenzel, and M. R. Anderson. 2000. Direct comparison of the impact of head-tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source. In: Proceedings of the 108th AES Convention. Paris, France.

[12] Elliott H. Berger, Ronald W. Kieper, and Dan Gauger. 2003. Hearing protection: Surpassing the limits to attenuation imposed by the bone-conduction pathways. Journal of the Acoustical Society of America 114, no. 4, pages 1955–1967.

[13] Meera M. Blatter, Denise A. Sumikawa, and Robert M. Greenberg. 1989. Earcons and Icons: Their Structure and Common Design Principles , 4:1,11 — 44. Human-Computer Interaction 4, no. 1, pages 11–44. URL `http://dx.doi.org/10.1207/s15327051hci0401_1`.

[14] M. Brandstein and D. Ward. 2001. Microphone arrays. Springer-Verlag, New York.

[15] Stephen A. Brewster. 1997. Navigating Telephone-Based Interfaces with Earcons. In: HCI 97: Proceedings of HCI on People and Computers XII, pages 39–56. Springer-Verlag, London, UK. ISBN 3-540-76172-1.

[16] Stepher A. Brewster, Peter C. Wright, and Alistari D. N. Edwards. 1995. Parallel earcons: Reducing the length of audio messages. International journal of human-computer studies 43, pages 153–175.

[17] D.S. Brungart, A.J. Kordik, C.S. Eades, and B.D. Simpson. 2003. The effect of microphone placement on localization accuracy with electronic pass-through

earplugs. In: Proceedings of IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, pages 149–152.

[18] W. R. D'Angelo, R. S. Bolia, P. J. Mishler, and L. J. Morris. 2001. Effects of CIC hearing aids on auditory localization by listeners with normal hearing. Journal of speech, language, and hearing research 44, no. 6, pages 1209–1214.

[19] J. H. Dibiase, H. F. Silrverman, and M. S. Brandstein. 2001. Robust localization in reverberant rooms. In: M. S. Brandstein and Eds. D. Ward (editors), Microphone arrays: Signal Processing Techniques and Applications, chapter 7, pages 131–154. Springler-Verlag.

[20] Christina Dicke, Jaka Sodnik, and Mark Billinhurst. 2007. Spatial auditory interfaces compared to visual interfaces for mobile use in a driving task. In: Proceedings of the Ninth International Conference on Enterprise Information Systems, pages 282–285. Funchal, Madeira, Portugal.

[21] Facebook. Accessed: May 13, 2009. http://www.facebook.com.

[22] Seppo Fagerlund. 2007. Bird Species Recognition Using Support Vector Machines. EURASIP Journal on Advances in Signal Processing Article ID 38637, no. doi:10.1155/2007/38637, page 8 pages.

[23] William W. Gaver. 1989. "The SonicFinder: An Interface That Uses Auditory Icons", Human-Computer Interaction, 4:1,67 — 94. Human-Computer Interaction 4, pages 67–94.

[24] Michael A. Gerzon. 1985. Ambisonics in multichannel broadcasting and video. Journal of the Audio Engineering Society 33, no. 11, pages 859–871.

[25] Samuel Gilman. 1983. Some factors affecting the performance of airline entertainment headsets. Journal of the Audio Engineering Society 31, no. 12, pages 914–920.

[26] David Gnewikow and Meredith Moss. 2006. Hearing aid outcomes with open- and closed-canal fitting. The Hearing Journal 59, no. 11, pages 66–72.

[27] Rebecca E. Grinter and Allison Woodruff. 2002. Ears and hair: what headsets will people wear? In: CHI '02: CHI '02 extended abstracts on Human factors in computing systems, pages 680–681. ACM, New York, NY, USA. ISBN 1-58113-454-1.

[28] Drexel Hallaway, Steven Feiner, and Tobias Höllerer. 2004. Bridging the gaps: Hybrid tracking for adaptive mobile augmented reality. Applied Artificial Intelligence Journal, Special Issue on AI in Mobile Systems 18, no. 6, pages 477–500.

[29] Dorte Hammershøi and Henrik Møller. 1996. Sound transmisssion to and within the human ear canal. Journal of the Acoustical Society of America 100, no. 1, pages 408–427.

[30] A. Härmä, M.F. McKinney, and J. Skowronek. 2005. Automatic surveillance of the acoustic activity in our living environment. In: Multimedia and Expo, 2005. ICME 2005. IEEE International Conference on.

[31] Aki Härmä, Julia Jakka, Miikka Tikander, Matti Karjalainen, Tapio Lokki, Jarmo Hiipakka, and Gaetan Lorho. 2004. Augmented Reality Audio for Mobile and Wearable Appliances. Journal of the Audio Engineering Society 52, no. 6, pages 618–639.

[32] Timo Hiekkanen, Aki Mäkivirta, and Matti Karjalainen. 2009. Virtualized listening tests for loudspeakers. Journal of the Audio Engineering Society 57, no. 4, pages 237–251.

[33] Marko Hiipakka, Miikka Tikander, and Matti Karjalainen. 2009. Modeling of external ear acoustics for insert headphone usage. In: Proceedings of the 126th AES Convention. Munich, Germany.

[34] Per Hiselius. 2005. Attenuation of Earplugs - Objective Predictions Compared to Subjective REAT Measurements. Acta Acustica united with Acustica 91, pages 764–770.

[35] Per Hiselius. 2006. Acoustics modeling of earplugs. Acta Acustica united with Acustica 92, pages 135–138.

[36] Bruce Hoff and Ronald Azuma. 2000. Autocalibration of an Electronic Compass in an Outdoor Augmented Reality System. In: Proceedings of IEEE and ACM International Symposium on Augmented Reality 2000, pages 159–164. Munich, Germany.

[37] Tobias Hans Höllerer. 2004. User interfaces for mobile augmented reality systems. Ph.D. thesis, Columbia University.

[38] J. D. Hood. 1962. Bone Conduction: A Review of the Present Position with Especial Reference to the Contributions of Dr. Georg von Békésy. The Journal of the Acoustical Society of America 34, no. 8, part 2, pages 1325–1332.

[39] T. S. Kapteyn, E. H. J. F. Boezeman, and A. M. Snel. 1983. Bone-conduction measurement and calibration using the cancellation method. Journal of the Acoustical Society of America 74, no. 4, pages 1297–1299.

[40] James M. Kates. 1988. A computer simulation of hearing aid response and the effects of ear canal size. Journal of the Acoustical Society of America 83, no. 5, pages 1952–1963.

[41] Shyam M. Khanna, Juergen Tonndorf, and Judith E. Gueller. 1976. Mechanical paramerers of hearing by bone conduction. Journal of the Acoustical Society of America 60, no. 1, pages 139–154.

[42] Andrew J. King, Jan W.H. Schnupp, and Timothy P. Doubell. 2001. The shape of ears to come: dynamic coding of auditory space. Trends in Cognitive Sciences 5, no. 6, pages 261 – 270.

URL http://www.sciencedirect.com/science/article/B6VH9-4379RWK-N/2/bbc00f3fd06f323898e14d9ba4290d9e.

[43] Robert W. Lindeman, Haruo Noma, and Paulo Gonçalves de Barros. 2007. Hear-Through and Mic-Through Augmented Reality: Using Bone Conduction to Display Spatialized Audio. In: Proceedings of the International Symposium On Mixed and Augmented Reality (ISMAR 2007), pages 173–176. Nara-Ken New Public Hall, Nara, Japan.

[44] Jack M. Loomis, Reginald G. Colledge, and Roberta L. Klatzky. 1998. Navigation system for the blind: Auditory display modes and guidance. Precence 7, no. 2, pages 193–203.

[45] Georgios Marentakis and Stepher Brewster. 2004. A Study on Gestural Interaction with a 3D Audio Display. In: Proceedings of MobileHCI2004, volume 3160, pages 180–191. Glasgow, Scotland.

[46] Natalia Marmasse and Chris Schmandt. 2000. Location-Aware Information Delivery with ComMotion, pages 361–370. Springer Berlin / Heidelberg.

[47] Russel L. Martin and Ken I. McAnally. 2001. Free-field equivalent localization of virtual audio. Journal of the Audio Engineering Society 49, no. 1/2, pages 14–22.

[48] Henrik Møller, Clemen Boje Jensen, Dorte Hammershøi, and Michael Friis Sørensen. 1995. Design Criteria for Headphones. Journal of the Audio Engineering Society 43, no. 4, pages 218–232.

[49] Henrik Møller, Mikael Friis Sørensen, Dorte Hammershøi, and Clemen Boje Jensen. 1995. Head-Related Transfer Functions of Human Subjects. Journal of the Audio Engineering Society 43, no. 5, pages 300–321.

[50] Elizabeth D. Mynatt, Maribeth Back, Roy Want, and Ron Frederick. 1997. Audio aura: light-weight audio augmented reality. In: UIST '97: Proceedings

of the 10th annual ACM symposium on User interface software and technology, pages 211–212. ACM, New York, NY, USA. ISBN 0-89791-881-9.

[51] Debra R. Parker Oliver, George Demiris, and Davina Porock. 2005. The usability of videophones for seniors and hospice providers: a brief report of two studies. Computers in Biology and Medicine 35, no. 9, pages 782 – 790. URL http://www.sciencedirect.com/science/article/B6T5N-4D4D132-3/2/fb34df6e4ed4785aea57ea313170ae37.

[52] Shinyoung Park, Akira Harada, and Hiroya Igarashi. 2006. Influences of personal preference on product usability. In: CHI '06: CHI '06 extended abstracts on Human factors in computing systems, pages 87–92. ACM, New York, NY, USA. ISBN 1-59593-298-4.

[53] C. A. Poldy. 1994. Loudspeaker and headphone handbook, chapter 12 Headphones. Focal Press.

[54] Ville Pulkki. 2001. Spatial Sound Generation and Perception by Amplitude Panning Techniques. Ph.D. thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland.

[55] Ville Pulkki. 2007. Spatial Sound Reproduction with Directional Audio Coding. Journal of the Audio Engineering Society 55, no. 6, pages 503–516.

[56] Ville Riikonen, Miikka Tikander, and Matti Karjalainen. 2008. An augmented reality audio mixer and equalizer. In: Proceedings of the 124th AES Convention. Amsterdam, The Netherlands.

[57] Niklas Röber and Maic Masuch. 2005. Playing audio-only games a compendium of interacting with virtual, auditory worlds. In: de Castell Suzanne and Jenson Jennifer (editors), Changing Views: Worlds in Play, page 8. University of Vancouver, Vancouver. URL http://www.digra.org/dl/display_html?chid=06276.30120.pdf.

[58] Nicoleta Roman and DeLiang Wang. 2008. Binaural tracking of multiple moving sources. In Press.

[59] Joseph Rozier, Karrie Karahalios, and Judith Donath. 2000. Hear & There: An augmented reality system of linked audio. In: International Conference for Audiotory Display (ICAD) 2000. Georgia Institute of Technology Atlanta, Georgia, USA. URL `http://www.icad.org/websiteV2.0/Conferences/ICAD2000/ICAD2000.html`.

[60] Nitin Sawhney and Chris Schmandt. 2000. Nomadic radio: speech and audio interaction for contextual messaging in nomadic environments. ACM Trans. Comput.-Hum. Interact. 7, no. 3, pages 353–383.

[61] Gerhard Schall, Erick Mendez, Ernst Kruijff, Eduardo Veas, Sebastian Junghanns, Bernhard Reitinger, and Dieter Schmalstieg. 2009. Handheld Augmented Reality for underground infrastructure visualization. Personal and Ubiquitous Computing 13, no. 4, pages 281–291. URL `http://dx.doi.org/10.1007/s00779-008-0204-5`.

[62] Barbara G. Shinn-Cunningham, Nathaniel I. Durlach, and Richard M. Held. 1998. Adapting to supernormal auditory localization cues. I. Bias and resolution. The Journal of the Acoustical Society of America 103, no. 6, pages 3656–3666. URL `http://link.aip.org/link/?JAS/103/3656/1`.

[63] Barbara G. Shinn-Cunningham, Nathaniel I. Durlach, and Richard M. Held. 1998. Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response. The Journal of the Acoustical Society of America 103, no. 6, pages 3667–3676. URL `http://link.aip.org/link/?JAS/103/3667/1`.

[64] Jaka Sodnik, Saso Tomazic, Raphael Grasset, Andreas Duenser, and Mark Billinghurst. 2006. Spatial sound localization in an augmented reality environment. In: OZCHI '06: Proceedings of the 20th conference of the computer-

human interaction special interest group (CHISIG) of Australia on Computer-human interaction: design: activities, artefacts and environments, pages 111–118. ACM Press, New York, NY, USA. ISBN 1-59593-545-2.

[65] Raymond M. Stanley and Bruce N. Walker. 2006. Lateralization of Sounds Using Bone-Conduction Headsets. In: Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES2006), pages 1571–1575. San Fransisco, CA.

[66] Walker Stanly. 2006. Toward Adapting Spatial Audio Displays for Use With Bone Conduction: The Cancellation of Bone-Conducted and Air-Conducted Sound Waves. Master's thesis, Georgia Institute of Technology.

[67] Stefan Stenfeld, Timothy Wild, Naohito Hato, and Richard L. Goode. 2003. Factors contributing to bone conduction: The outer ear. Journal of the Acoustical Society of America 113, no. 2, pages 902–913.

[68] Ryuhei Tenmoku, Masayuki Kanbara, and Naokazu Yokoya. 2003. A Wearable Augmented Reality System for Navigation Using Positioning Infrastructures and a Pedometer. In: Proceedings of the IEEE/ACM International Symposium on Mixed and Augmented Reality, ISMAR '03. The National Center of Sciences, Tokyo, Japan.

[69] Günther Theile. 1986. On the standardization of the frequency response of high-quality studio headphones. Journal of the Audio Engineering Society 34, no. 12, pages 956–969.

[70] F. Thomas and L. Ros. 2005. Revisiting trilateration for robot localization. Robotics, IEEE Transactions on 21, no. 1, pages 93–101.

[71] Miikka Tikander. 2007. Modeling the attenuation of a loosely-fit insert headphone for augmented reality audio. In: Proceedings of the AES 30th International Conference. Saariselkä, Finland.

[72] Miikka Tikander, Aki Härmä, and Matti Karjalainen. 2004. Acoustic positioning and head tracking based on binaural signals. In: Presented at the 116th AES convention. Berlin, Germany.

[73] Twitter. Accessed: May 13, 2009. http://twitter.com/.

[74] Frederic A. Tyszka and David P. Goldstein. 1975. Interaural phase and amplitude relationships of bone-conduction signals. Journal of the Acoustical Society of America 57, no. 1, pages 200–206.

[75] Maria L. M. Vargas and Sven Anderson. 2003. Combining speech and earcons to assist menu navigation. In: Proceedings of the 2003 Internatioanal Conference on Auditory Display. Boston, MA, USA.

[76] Sampo Vesa and Aki Härmä. 2005. Automatic estimation of reverberation time from binaural signals. In: Proceedings of the International Conference on Acoustics, Speech, and Signal Processing (ICASSP '05), volume 3, pages 281–284. Philadelphia, USA.

[77] Sampo Vesa and Tapio Lokki. 2005. An eyes-free user interface controlled by finger snaps. In: Proceedings of the 8th International Conference on Digital Audio Effects (DAFx'05), pages 262–265. Madrid, Spain.

[78] Sampo Vesa and Tapio Lokki. 2006. Detection of room reflections from a binaural room impulse response. In: Proceedings of the 9th International Conference on Digital Audio Effects (DAFx'06), pages 215–220. Montreal,Canada.

[79] K. Vickery. 1998. Acoustic positioning systems. A practical overview of current systems. In: Autonomous Underwater Vehicles, 1998. AUV'98. Proceedings Of The 1998 Workshop on, pages 5–17.

[80] Edgar Villchur. 1969. Free-field calibration of earphones. Journal of the Acoustical Society of America 46, no. 6 (Part 2), pages 1527–1534.

[81] Michael Vorländer. 2000. Acoustic load on the ear caused by headphones. The Journal of the Acoustical Society of America 107, no. 4, pages 2082–2088. URL http://link.aip.org/link/?JAS/107/2082/1.

[82] A. Walker, S.A. Brewster, D. McGookin, and A Ng. 2001. Diary in the sky: A spatial audio display for a mobile calendar. In: In: 15th Annual Conference of the British HCI Group. Lille, France.

[83] Bruce N. Walker and Raymond M. Stanley. 2005. Thresholds of Audibility for Bone-Conduction Headsets. In: Proceedings of the International Conference on Auditory Display, pages 218–222.

[84] Bruce N. Walker, Raymond M. Stanley, Nandini Iyer, Brian D. Simpron, and Douglas S. Brungart. 2005. Evaluation of Bone-Conduction Headests for Use in Multitalker Communication Environments. In: Proceedings of the Annual Meeting of the Human Factors and Ergonomics Society (HFES2005), pages 1615–1619. Orlando, FL.

[85] Suya You, Ulrich Neumann, and Ronald Azuma. 1999. Orientation Tracking for Outdoor Augmented Reality Registration. IEEE Computer Graphics and Applications 19, no. 6, pages 36–42.

[86] Vasileios Zeimpekis, Gerge M. Giaglis, and George Lekakos. 2003. A Taxonomy of Indoor and Outdoor Positioning Techniques for Mobile Location Services. SIGecom Exch. 3, no. 4, pages 19–27. URL http://doi.acm.org/10.1145/844351.844355.