



Audio Engineering Society Convention Paper 5768

Presented at the 114th Convention
2003 March 22–25 Amsterdam, The Netherlands

This convention paper has been reproduced from the author's advance manuscript, without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. Additional papers may be obtained by sending request and remittance to Audio Engineering Society, 60 East 42nd Street, New York, New York 10165-2520, USA; also see www.aes.org. All rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Techniques and applications of wearable augmented reality audio

Aki Härmä¹, Julia Jakka¹, Miikka Tikander¹, Matti Karjalainen¹, Tapio Lokki², Heli Nironen², Sampo Vesa²

¹Laboratory of Acoustics and Audio Signal Processing, Helsinki University of Technology, P.O. Box 3000, 02015, HUT, Finland

²Telecommunications Software and Multimedia Laboratory, Helsinki University of Technology, P.O. Box 5400, 02015, HUT, Finland

Correspondence should be addressed to Aki Härmä (Aki.Harma@hut.fi)

ABSTRACT

The concept of augmented reality audio characterizes techniques where real sound environment is extended with virtual auditory environments and communications scenarios. This article introduces a framework for Wearable Augmented Reality Audio (WARA) based on a specific headset configuration and a real-time audio software system. We will review relevant literature and aim at identifying most potential application scenarios for WARA. Listening test results with a prototype system will be presented.

1 INTRODUCTION

The era of wearable audio appliances started with the introduction of a portable cassette player two decades ago. The development of digital technology led to portable CD players and finally to fully digital mp3 players. Digital cellular phones have been around for ten years. While speech communications is the main application, many manufacturers have recently integrated a digital audio player to a phone to enable high-quality audio playback. However, the basic application scenario for wideband audio is still the same as in early Walkmans. Another aspect standing with the aging population in developed countries is that the number of users for hearing aid devices is constantly increasing. With digital technology the quality of hearing aid devices has significantly improved while the prices are dropping and thus preparing the way for even higher number of users. Yet another application that relates to this is personal active hearing protectors.

We may have multiple wearable audio appliances but only one pair of ears. At some point it may make sense to integrate all those functions into the same physical device. Mechanical and electrical integration is already feasible. However, in application scenarios there are many interesting new possibilities and problems to explore. Also, the progress in speech and audio technology, computing, and communications predict introduction of completely new types of intelligent and interactive audio and speech applications. For example, auditory displays which can provide a user different types of information in the form of spatialized sound events have been introduced.

We consider a device which a user could be wearing at all times. It would resemble portable audio players in some respects and also provide speech and audio communications services, e.g., over a wireless network. But, at the same time, it would also make it possible for a user to hear and interact with the real acoustic environment in a natural way. Thus it would facilitate or even make easier ordinary speech communications with other people, enable safe navigation in traffic, and operation of machines where acoustic feedback is important. In addition, there would be a large number of new functions which provide information and communications channels which are not available in a natural acoustic environment or in current appliances.

The possibility to hear the natural acoustic environment

around a user differentiates the concept of augmented reality audio from the traditional concept of a virtual audio environment where a user is typically immersed into a completely synthetic acoustic environment. The proposed system for WARA requires specific transducer systems and auralization techniques. In the prototype system introduced in this article the transducer configuration is based on a headset where miniature microphones are integrated into earphone elements in both ears. When microphone sounds are routed directly to earphones, a user can perceive a representation of the real acoustic environment. Since the experience may differ from the open-ear case we call this representation a *pseudo-acoustic* environment. It has been demonstrated that users can adapt to a modified binaural representation [1]. Virtual and synthetic sound events, such as talk of a remote user, music, or audio markers are superimposed onto the pseudo-acoustic sound environment in a device which may be called an augmentation mixer. At one extreme, virtual sounds can be combined with the pseudo-acoustic signals in the augmentation mixer in such a way that a user may not be able to determine which sound sources are local and which ones are artificially rendered by means of digital signal processing. Preliminary listening tests show that this is relatively easy to achieve in using personalized in-situ HRIRs in the synthesis of virtual sounds.

Real-time implementation of WARA system requires low latency for audio and seamless integration of audio streams, signal processing algorithms and network communications. We will introduce a modular and flexible software architecture developed for testing of different application scenarios.

One class of applications is based on localized audio messages. For example, we may consider *auditory Post-It* application where a user can leave and receive audio messages related to certain places or objects in our environment. We review the technology for detecting the location and orientation of a user in those of applications and report recent experimental results. In speech communication applications the use of binaural microphones makes it possible to explore new ways of communication between people.

We start with an introductory section where we put the WARA concept into a wider context and define the central parts of the proposed framework. Then we review the literature and give an organized representation of pre-

vious works and ideas in related application fields and hardware solutions available at the time of writing this paper. In Section 4 and 5 we introduce a software platform developed for testing of different techniques and applications.

2 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. Augmented reality audio (or augmented audio reality) 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. At one extreme an augmented reality audio system should pass a test which is closely related to the classical *Turing test* for artificial intelligence [2]. That is, if a listener is unable to determine whether a sound source is a part of the real or a virtual audio environment, the system implements a subjectively perfect augmentation of listener's acoustic environment. At the other extreme, virtual auditory scenes could be rendered in high quality such that they are easily separable from real ones by their characteristics that are not possible in normal acoustic environments.

In the current paper the focus is in developing techniques for future wearable applications. Hence, it is clear that transducers used for producing virtual sounds must be wearable. In practice, headphones or earphones are most viable alternatives. Headphones have been used successfully in many virtual reality applications reviewed in Section 3. Figs. 1 a and b illustrate a user in a real acoustic environment and with a headphone-based virtual acoustic system where sound environment is created using a computer.

2.1 Headphone problems in virtual spatial audio

Headphone auralization often leads to a perceived effect of having the virtual source localized inside listener's head. This is usually called *intracranial*, or inside-the-head locatedness (IHL) [3]. Spatial localization of sound sources which are perceived to be inside listener's head is termed *lateralization*. It has been demonstrated that a listener can make a clear distinction in headphone listening between *localized* (that is, sounds outside-of-head) and lateralized sound sources and that these two types can coexist in listener's experience [4].

The effect of a lateralized sound in headphone listening can be produced using amplitude and delay differences in

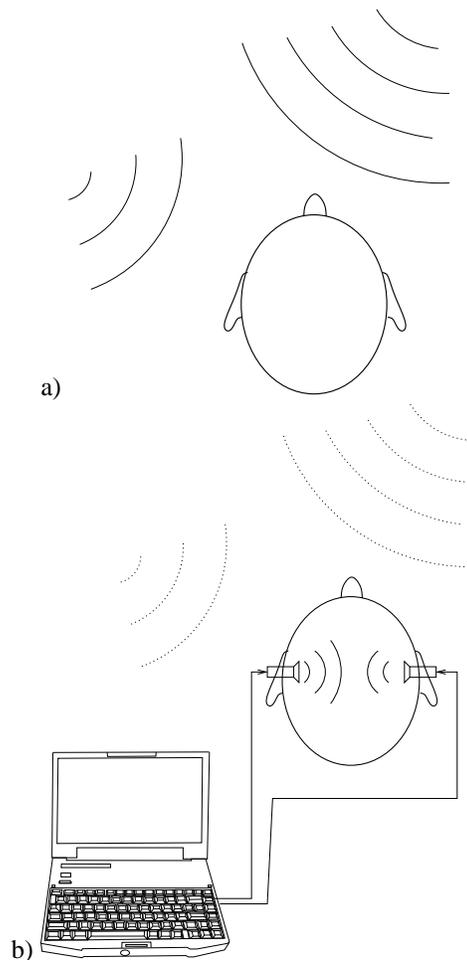


Fig. 1: A listener in a) real and b) virtual environment.

two headphone channels corresponding to each source. In order to make a sound source *externalized*, more sophisticated binaural techniques are needed [5]. In particular, we may list the following aspects

1. Spectrum differences in the two ear signals due to *head-related transfer functions*, HRTF's. In laboratory environment *personalized* HRTF's can be used to produce a realistic illusion of an externalized source [6]. However, there is a great variability in HRTF's among subjects.
2. Acoustic cues such as the amount of reverberation in virtual sound. It has been demonstrated that the use of artificial reverberation can help in forming an externalized sound image in headphone listening

[7].

3. Dynamic cues related to head turning and other movements of a listener or a source. The virtual source should be stable in relation to the environment. In headphone listening this requires that the auralization processor is controlled by information on the position and orientation of listener's head, see review on head-tracking techniques in Section 6.
4. Multimodality aspects such as connection of a sound event to a visible real-world object.
5. User's expectations on the performance of the system also affect externalization and localization of virtual sound sources, see, e.g., [8].

2.2 The proposed transducer system

In this study, we focus on a rather specific type of microphone-earphone systems. There are many other alternatives but this was chosen as a starting point because it seems to have many beneficial properties and it directly facilitates testing of rather unconventional applications and services. This is a microphone-earphone system where microphones are mounted directly on earplug-type earphone elements and are therefore placed very close to the ears. Ideally, the whole system could fit in user's ears and there would be no additional wires. The wearability of this device would be the same as for a hearing aid device. Earplug-type earphones have a low power consumption which results in extended operating times and low weight of the device. For testing we constructed two different types of headsets. Attenuation of direct sound in model I measured in an anechoic chamber with an audiometer is plotted in Fig. 2. In model I, the earphones are of earplug-type and provide 10-30 dB attenuation of direct sound. In model II, the headphones are traditional open earphones placed at the entrance of the ear canal and providing only 1-5 decibel attenuation. The right ear element of the model II is shown in Fig. 3.

The use of earplug-type headphones makes it possible to control the signals entering listener's ears more accurately than, e.g., with open headphones. For example, one may mute or amplify sounds selectively if the direct acoustic propagation of sound into the ears is suppressed by the blocking of the ear canal by the device. The attenuation of external sounds is one of the parameters which need to be studied in comparing different transducer systems.

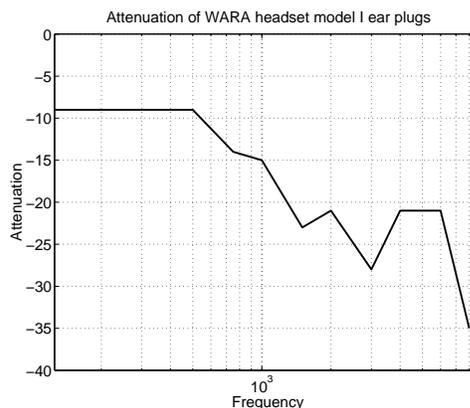


Fig. 2: Attenuation in headset model I (earplug).

Microphones are located in the ears or very close to the ears at both sides. When microphone signals are routed directly to earphones the system exposes a user to a 'binaural' representation of the real acoustic environment around the user. However, in practice it is almost impossible to position and tune transducers so that signals entering listener's ears are identical to those in the open-ear case. Hence, the produced spatial impression corresponding to the real acoustic environment is also altered. To make a distinction between the physically real acoustic environment and its electronic representation in user's ears we call the latter one a pseudo-acoustic environment. This is illustrated in Fig. 4 a, where a user is wearing a microphone-earphone system.

2.3 Pseudo-acoustic environment

The pseudo-acoustic environment is a modified representation of the real acoustic environment around a user. In many applications it is most convenient to try to make the pseudo-acoustic environment as identical to the real environment as possible. In principle, this could be achieved by means of digital filtering of microphone signals. Equalization filters could be estimated in a measurement where a probe microphone was inserted into the ear canal. However, this is difficult and would probably lead to highly individualized filters specific for some particular piece of equipment. Therefore, in the current phase, we only try to control the signal level and do some coarse equalization of signals to make the *pseudo* environment sound as natural as possible. Accordingly, some difference is expected to remain in the user's spatial impression. However, it has been demonstrated in many experiments that listeners can adapt to atypical [1] or *su-*

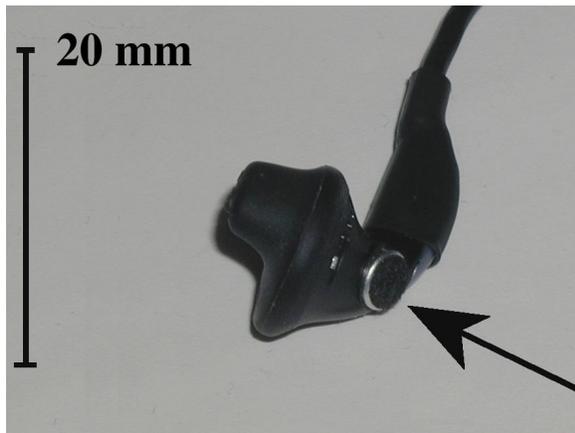


Fig. 3: Model II headset element with an open-type earphone (constructed of a Sony MDR-ED268LP earphone and an electret microphone element). Position of the microphone element is indicated by the arrow.

pernormal [9] binaural inputs. Therefore, we may expect that a user could adapt to the use of the proposed transducer system, too.

The proposed model could also be used to produce a modified representation of reality which could be advantageous, more convenient, or just an entertaining feature for a user. For example, in some cases the system would provide hearing protection, hearing aid functions, noise reduction, spatial filtering, or emphasize important sound signals such as alarm and warning sounds.

2.4 Virtual sound environment

A virtual sound environment is created using traditional techniques of auralization [10, 11, 12, 13]. This mainly involves binaural filtering using HRTF filters. However, this may not be sufficient for good results in externalization of sound sources especially in a reverberant pseudo-acoustic environment. Therefore it is probably necessary to bring some early reflections and reverberation also to the virtual sound environment to make it match better with the local environment. It is currently an open research issue how this should actually be done and therefore many different approaches have to be tested.

In some applications the rendered virtual sound environment should be independent of the position and rotation of user's head. The sound source should be *localized* and often somehow connected to the real environment around the user. In those applications some system for finding

the position and orientation of a user is needed. Several alternative techniques for head-tracking are reviewed in Section 6.

2.5 Augmented reality audio environment

An *augmented audio environment* is produced by superimposing *virtual sound environment* onto the *pseudo-acoustic environment*. First, the local environment should be delivered to a user in such a way that its loudness and binaural properties are acceptable for the user. Secondly, virtual sound environment should be mixed with the local environment carefully to produce a coherent perception for a user. A goal is to find the best *mixing rules* for local and virtual environments leading to a meaningful fusion of these two main components of augmented reality audio. In Fig. 4b the mixing of the pseudo-acoustic and virtual audio environments is performed in a device which is called augmented reality audio (ARA) mixer.

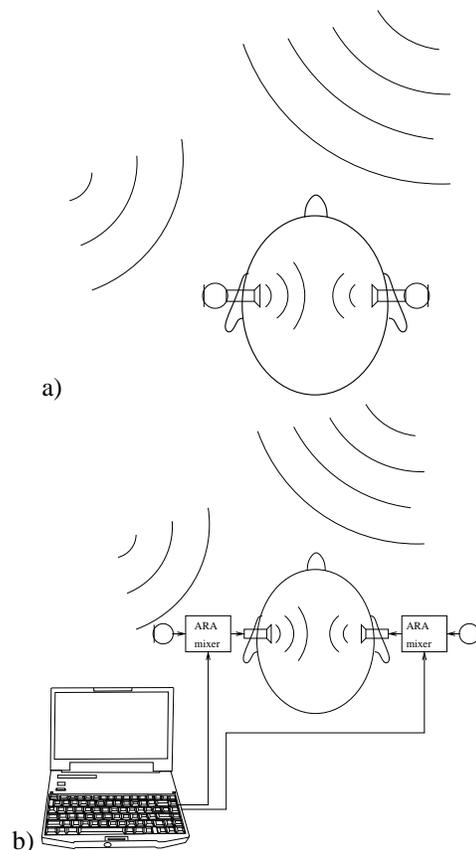


Fig. 4: A listener in a) pseudo-acoustic and b) augmented environment.

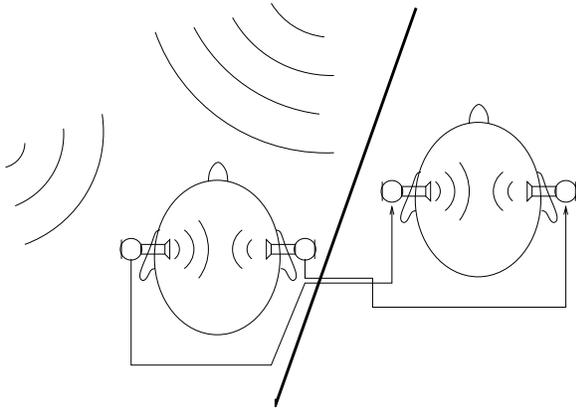


Fig. 5: A listener in an augmented environment and another user experiencing telepresence based on a real-time binaural recording.

The proposed scheme makes possible a number of different application ideas. One of them is illustrated in Fig. 5. The person on the left hand side is using the WARA system introduced earlier and the person on the right at a remote location can experience telepresence based on direct transmission of first user's pseudo-acoustic environment. For the user on the right, the ARA mixer would combine the *local* pseudo-acoustic environment with *remote* pseudo-acoustic environment to produce a specific type of augmented audio reality experience.

2.6 Listening test

A preliminary listening test was carried out in order to evaluate the level of performance of the WARA system in virtually imitating localized pseudo-acoustic sound. The test was carried out in a standardized listening room and authors of this paper were used as test subjects. The setup consisted of a head-supporting chair for the test subject wearing the WARA headset, and a loudspeaker not visible to the subject. A similar test setup has been used earlier in studies on externalization of virtual sources, e.g., in [6]. The listening test system is illustrated in Fig. 6

In the beginning of the test session, an impulse response from the loudspeaker to the WARA headset microphones in the test subject's ears was measured. The binaural impulse response contains both the subject's individual head-related impulse response in relation to the location of the loudspeaker, as well as the room response. Two types of test signals were then played for the test subject to compare. The pseudo-acoustic signals were acousti-

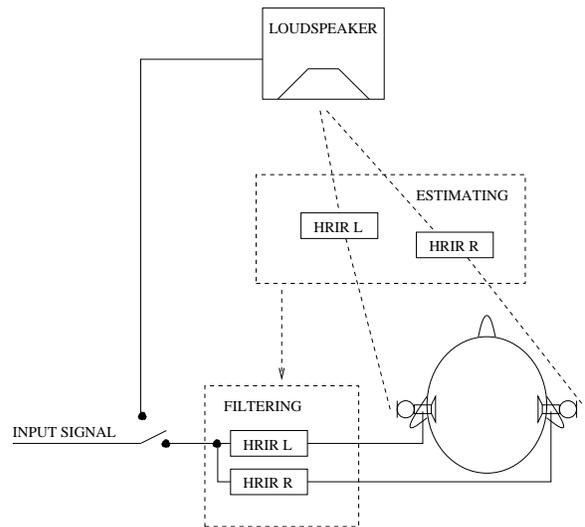


Fig. 6: The listening test configuration. Head-related Room Impulse Responses (HRIR) were measured for each listener. The HRIRs were then used to synthesize test signals for headset playback. In a listening test, test signals were switched between loudspeaker and ear-phone playback.

cally dry signals played from the loudspeaker. The virtual test signals were the same signals filtered with the measured binaural response in order to auralize them. The test was a two-alternative forced choice test, where the subject was asked which one of the signals is played from the loudspeaker.

As both types of the test signals are perceived through the earphones, hypothetically no difference in the signals should be audible. On the other hand, a perceivable difference may occur due to the direct sound leaking into the ear past the headset system in the case of loudspeaker playback, and also the low frequency partials being conducted to the ear by bone. In particular, the model II headset attenuated direct sound leaking into the ear by only a few decibels. In the listening test, the level of the pseudoacoustic environment in listeners ears was 3-5 dB higher than would be without the headset. The listening test results show that indeed pseudo-acoustic sound can be almost indistinguishably imitated. The test subjects were able to recognize or guess the source of the sound correctly only in 68% of the cases (pure guessing gives 50 %). The sources of the samples with broader frequency band were easier to distinguish, whereas dis-

tinguishing the sources of simple instrument sounds or speech was almost impossible. It seems that the pseudoacoustic sound masks the leaking direct sound very efficiently.

The test was carried out using the WARA model II headset. Similar results are obtained with the headset model I with a limited number of listeners. A more extensive listening test on mixing of pseudo-acoustic and virtual sound environments is currently being prepared.

3 TAXONOMY AND OVERVIEW OF PREVIOUS APPLICATIONS

Different types of applications can be categorized in many ways. Some cases are clearly examples of *communications* while in some cases they can be seen as *information services*. Another meaningful division is to consider a particular application to represent *human-to-human* or *human-to-machine* communications. Another important division is in the way how virtual audio events are connected to the real environment. In some cases created virtual events are *localized* and in other cases *freely-floating*. In the following, ideas and applications have been categorized into three different classes.

3.1 Speech communications

It can be expected that speech communications will remain as one of the most important applications for wearable augmented reality audio systems. However, it may take many different forms. As a communications application Figure 1b would show the case where a remote talker is represented to a user using techniques of binaural auralization. Fig. 4 b shows the case of augmented reality audio where the virtual source is mixed with the pseudo-acoustic environment around the user. Conceptually this can be interpreted as a situation where the remote talker has been rendered to user's local environment. Figure 5 illustrates a quite different scenario where the user on the right hand side is virtually transported to the location of the first talker. This is called telepresence.

The idea of telepresence was approached already in 1933, when binaural hearing was examined by letting test persons with two receivers listen to sound examples through microphones in a dummy head's ears, the dummy head being in another room [14].

Cohen et al. introduced a system where a robot puppet was controlled from a distance by a pilot wearing effectors corresponding to the puppet's sensors, so that the puppet could venture into hazardous environments

like fires or toxic waste dumps [15]. The relationship between acoustic telepresence and virtual reality presentations was examined, considering virtual reality as the perception of the sound field by the dummy head, yielded by HRTF filtering. Also the user dynamically selecting the transfer functions and thus placing the sound sources himself was experimented. No head-tracking was used.

Visiphone [16] is an "augmented home" system, a telephone augmented with graphics, a system that can be used to enable a casual hands-free limitedly mobile conversation over distance with a visual reminder, feedback system and a focal point for directing speech. In some sense the idea could be defined as an audio version of internet relay chat.

Another speech communications application is teleconferencing. An application of teleconferencing, a spatialized stereo audio conferencing system was designed by J. Virolainen [17]. The advance of being able to virtually localize the other attendees around the user helps the user to separate the different speakers making the conference more natural and easier to lead among multiple attendees. The naturalness and accuracy of the perception of directions are significantly increased when adding head-tracking to the system. It is also possible to extend the use of the application to wearable augmented reality teleconferencing.

3.2 Localized acoustic events connected to real world objects

The idea here is to immerse in one's normal audio scene additional audio information of the real environment, e.g., alerting sounds, instructions, even advertisements. The point is in connecting them to real world objects by auditorily localizing them. Superimposing synthesized sound on top of the real audio scene instead of isolating one from the real world by building a virtual audio reality was originally proposed by M. Krueger [18].

The LISTEN project [19] reports on augmenting the physical environment through a dynamic soundscape, which users experience over motion-tracked wireless headphones. The application experiment promotes a type of information system for intuitive navigation of visually dominated exhibition space, in this experiment a museum. The augmenting part of the application consists of speech, music, and sound effects creating a dynamic soundscape including exhibit-related information as well as "the right atmosphere" immersed individually in the visitor's auditory scene based on his/her spa-

tial behaviour and also on preliminarily entered information. The experiment includes a pedagogical as well as an artistic approach.

The idea of an automated tour guide was already proposed by Bederson [20] to replace the taped tour guides, which were found even obtrusive considering normal socializing. Bederson's prototype application included a random access digital audio source, a microprocessor to control the audio source, an infrared receiver to tell the processor the location of the viewer, and an infrared transmitter in the ceiling above each object to tell the system to play the pre-recorded descriptions of the piece, and also to stop the playing when the subject moved away.

A spatial auditory display was evaluated by air crews on simulated night mission flights [21]. The ordinary visual display was compared to one augmented with spatial auditory display. The results indicate that an improved auditory system would improve navigation.

An aiding navigation system for visually impaired is considered in [22], where the person's visual environment is registered by a camera and transformed to a sound scene providing information of obstacles or possible danger. The equipment used is a camera, a portable computer, and a pair of earphones.

A method to teach visually impaired persons orientation and mobility skills has been described by Inman [23], where simulated environments are used to train sound identification, localization and tracking sound sources. The simulation was implemented on a performance platform of a Pentium-based computer, headphones, a head-tracker and a joystick. As the student moved his or her head, the virtual environment sound source or sources changed accordingly. The different training tasks included identifying and reacting to the simulation sound sources in a particular way, completing an action like crossing a street in the virtual environment and completing an action by getting audio feedback from the student's own actions like walking along a line. The environment was also displayed on a screen so that the trainer can follow the student's actions.

The KnowWhere System is a method to present geographic information to blind people [24] by using a touchable virtual map surface recorded by a camera and transferring the features touched in the map to signals describing the objects.

Guided by Voices [25] is a game application that uses audio augmented reality system to provide location-triggered context-related playback sounds that inform the player. The sounds depend on the current location and also the history of the locations of an individual player. The system is based on wearable computers and an RF based location system with transmitters placed in the environment.

Another game based on auditory display system is Sleuth [26], a game where the player is set in the middle of a crime scene where he is supposed to use the information given by the auditory clues and determine the course of the crime and find out the one guilty.

3.3 Freely-floating acoustic events

The concept of freely-floating acoustic events, compared to localized events, are events that are not connected to objects in the subject's physical environment. Typically this means that the only anchor point relative to which the event is localized is user's head. Potential applications are information services such as news, calendar effects, or announcements, or many different forms of entertainment such as music listening.

A parade example of a freely-floating spatial audio event is in the DateBook application introduced in [27]. Here the calendar application of a PDA device was remapped to a spatial audio display of calendar events. Calendar entries at different times were rendered around the user so that noon appeared at the front of the user, 3 pm at the right and 6pm behind the user.

The Nomadic Radio project [28, 29] presents an idea of a wearable audio platform that provides access to all nowadays normal desktop messaging and information services as well as an awareness of people and background events situated elsewhere in a location of individual interest. The audio messaging and information services, such as email, voice mail, news broadcasts and weather reports, are downloaded to the device throughout the day and delivered to the user as synthesized speech. The messages are localized around the user's head based on their time of arrival. Different categories of messages can also be assorted to different levels of urgency or interest, and localized to different distances away from the head. The awareness of a certain remote location is based on the Cocktail Party Effect [30], meaning that one is able to monitor several audio streams simultaneously, selectively focusing on one and leaving the rest in the background. This is even easier if the streams are segregated

by localizing them away from each other. The device chosen as the Nomadic radio is a wearable shoulder-mounted audio-only device. The messages are broadcasted as spatial audio stream and synthesized speech. The navigation and controlling of the system is done by using a set of speech commands.

MacIntyre et al. [31] considered a combination of auditory and visual displays to provide both foreground interaction and background information and on the other hand both public and personal information. The focus is in developing new user interface paradigms and software, and creating testable prototype applications. The current audio part of the system is Audio Aura [32]. Audio Aura is a distributed system that employs a wearable active badge to transmit the individualized information to wireless headphones. The goal is to provide the user with auditory cues on other people's physical actions in the workplace, such as information on how long ago a certain person has left a room or where he is at the moment. The system detects the traffic and updates a database. Certain changes in the database result the auditory cues to be sent.

One aspect of localizing acoustic events is sonification. It is mostly understood as an additional, supporting dimension of visual interfaces to help navigation or performing tasks. Sonification as a topic has been studied extensively. The concept in relation to a prototype computer graphics extension has been studied by V. Salvador et al. [33] and an experiment with mobile phone interface was carried out by S. Helle et al. [34]. In the computer graphics system study it was found that there are even more ways to implement sonification in the system than was predicted. In the mobile phone experiment the sonification was found both irritating and also helpful in some occasions.

3.4 Other topics

In relation to wearable augmented audio equipment there are some interesting new aspects concerning hearing aid design. To reduce noise amplified along with the desired signal (usually speech), some new techniques have been introduced. Perceptual Time-Frequency Subtraction algorithm presented by Min Li et al. [35] is developed to simulate the masking phenomena and thereby enhance the speech signal, and it is meant to be applicable both in digital hearing aids and portable communication systems.

Individualizing spatialized audio displays is researched

by P. Runkle et al. [36] and a system of active sensory tuning, AST, is developed, where subjective preferences are possible to set.

4 THE WARA SOFTWARE SYSTEM

A generic block diagram of the WARA system sketched in Section 2 is shown in Fig. 7. The signals entering user's ears are composed of pseudo-acoustic input signals captured with microphones located close to the ears, and virtual sounds which may be speech signals from remote talkers or some other signals, such as recorded or synthesized announcements, advertisements, instructions, warnings, or music. Mixing of the two compounds is performed in the Augmented Reality Audio (ARA) mixer. The network shown above the head in Fig. 7 is used to produce the output signal of the system. Microphone signals are recorded and some preprocessing may be applied to produce a monophonic signal which is, in a communications application, transmitted to another user. Alternatively, the binaurally recorded signal (remote pseudo-acoustic signal for the other user, see, Fig. 5) could be transmitted in its original form. Both options have been implemented in the current software system.

The WARA software system is written in C++ and is (almost) completely built over the Mustajuuri framework. The Mustajuuri is a plugin-based real-time audio signal processing software package which is designed for low-latency performance [37]. The different configurations required by applications are constructed by adding plugins into a mixer console window (Figure 8) of the Mustajuuri program. The Mustajuuri plugin system allows different configurations to be built in a very quick and flexible way.

5 WARA APPLICATIONS

Different application schemes have been tried using the plugins of the WARA software system as building blocks. In this section these applications are overviewed and discussed.

5.1 Binaural telephone

The basic application of the WARA system is a binaural telephone. It is a two-way voice-over-IP communication scheme with binaural signals that WARA transducers provide. Binaural telephone enables a kind of telepresence instead of normal telephone, since the whole 3-D soundscape is transmitted to the receiver end. The binaural signals are transmitted via UDP protocol in its orig-

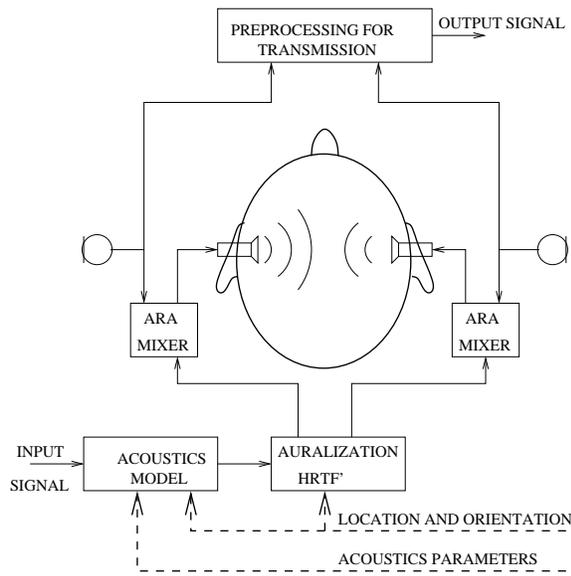


Fig. 7: A generic system diagram applicable for the majority of augmented reality audio applications studied in this report.

inal form to the remote end of communication. In current implementation no compression is applied to binaural signals, but for more efficient transfer some compression could be utilized.

In informal usability testing it was found that implemented binaural telephone really gives a convincing feeling of telepresence. The remote end soundscape is naturally transmitted to the other user. It is easy to get confused on which people actually are at which room, if there are other people besides the ones wearing the headsets in both ends. However, when utilizing the system in a meeting (only one user wearing WARA headset) an obvious shortcoming was discovered: the user at the remote end gets frustrated when he or she cannot talk to the other persons not wearing the headset at the other end of communication. Another problem in such a multiuser situation is that the volume level of the local speaker tends to be much louder than the level of the other speakers in the room.

5.2 Speech communications with headtracked auralization

When applying the WARA system as a binaural telephone the remote talker is localized inside the user's head. This does not sound very natural and it would



Fig. 8: Mustajuuri mixer console with two-way speech communications application.

be better if the remote talker can be positioned to a certain place in user's auditory environment. To position a remote talker, a speech communication application with headtracked auralization has been implemented. This application is similar to the binaural telephone, but it contains also plugins for head-tracked or non-head-tracked auralization for the incoming audio signal. The binaural signal is transmitted via UDP protocol in its original form and then converted to a mono signal and finally auralized with HRTFs. An example of such an application, implemented with Mustajuuri [37] is illustrated in Fig. 8.

The remote talker can be positioned outside the user's head by means of room acoustic modeling and HRTF filtering. Room acoustic modeling would consist of the synthesis of a number of early reflections and late reverberation.

In the current version, the WARA auralization has been implemented by applying a well-known direct sound, early reflections and statistical late reverberation concept (see, e.g., [10]). For auralized sound source (a remote talker) the direct sound and 14 early reflections, six for the first-order and eight for the lateral plane second-order early reflections are calculated from a simple shoe-box

room model with user-adjustable wall, floor, and ceiling locations. Each of these reflections is panned to correct direction using FIR filters, the responses of which have been matched to the minimum-phased HRTF functions. The applied filter orders for each of the three virtual sound source groups (direct sound, 1st and 2nd order reflections) can be adjusted individually and the computational burden can thus be reduced [38]. Informal listening tests suggest that the early reflections can be rendered with less filter taps than the direct sound without degrading the spatial impression.

In addition to the HRTF filtering, each early reflection is filtered with material absorption. Such filtering smooths out the comb-filter-like coloration caused by successive summing of an identical sound source - the fact that $1/r$ -law attenuation is implemented and the echoes come from different directions means that the summed sounds are not exactly identical, but close enough for some unnatural coloration to take place. In current version of the WARA auralization the material filters are first order IIR filters with coefficients calculated from given absorption coefficients at octave bands [38]. The most important thing with the material filtering is just to get somewhat realistic low-pass filtering for the reflections. The early reflections and their parameters are controlled with a plugin illustrated in Fig. 9.

The early reflections, although they are computed from a simplified geometry, help in externalization of the virtual sound sources, but the auralized sound is still somewhat unnatural. For even better results, some diffuse late reverberation is added to match the length of the reverberation with the reverberation of the pseudo-acoustic environment. The applied reverberation algorithm [39] is a variant of feedback delay network (FDN) algorithm. The parameters for the artificial reverberation (reverberation time etc.) are chosen manually by analysing the reverberation of the pseudo-acoustic environment. A more sophisticated approach would be to estimate parameters automatically from the binaural microphone signals (see section 6.4).

In cases where virtual sources are *localized* in the environment both room modeling and binaural synthesis need additional information about the *orientation and location* of the user. Orientation and location information of a user can be estimated using some head-tracking device. A limitation in many current technologies (see Section 6) is the need of an external device placed in the en-

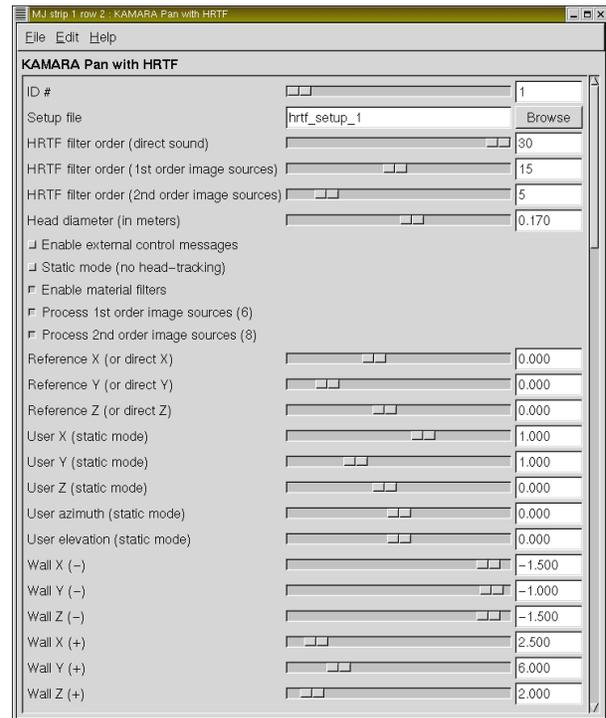


Fig. 9: Control window of the HRTF auralize plugin.

vironment. Those can be used only in a close range and are therefore basically limited to applications where an augmented sound source is localized to some real world object which could host also a *transmitter* of a head-tracking system.

5.3 Auditory Post-it application

The WARA framework enables also different location-based applications. One such application is an auditory Post-it, in which short messages can be attached to certain objects or messages can be left to certain places.

With current implementation of auditory Post-it, it is possible to leave and receive five messages. However, with small adjustments to the code it would be possible to compile versions with any number of messages. A message is left by recording it to one of the files found from the common soundfile directory. The coordinates of the message(s) are given in a text file. Messages are received by selecting command *Get new messages*, see Fig. 10. After this by selecting *Play all messages once*, *Play all messages repeatedly* or *Play selected message* it is possible to listen received messages. If the last one: *Play selected message* is chosen, user has also to select a mes-

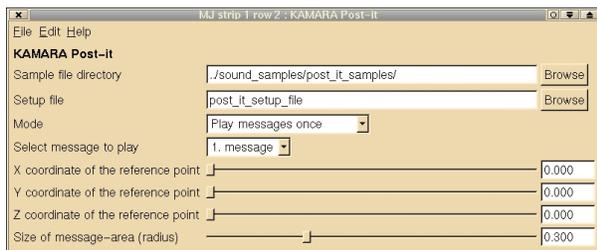


Fig. 10: Auditory Post-it application

sage to play. All these commands are controlled using a GUI (Fig. 10). It is also possible to change the coordinates of the reference point (The coordinates of messages given in the text file are referenced to this point) and the size of the message area (radius). The localization of the user is done by a head-tracker which sends the coordinates of the user to the application. When user, wearing a detector, is inside a message area, the message is played if selected command allows that.

The current version of the auditory Post-it application works well when the messages are located not more than 1-1.5 m from the head-tracker. However, with better tracker it should be possible to detect messages from a larger area. The next improvement to the auditory Post-it application will be the implementation of the possibility to leave a message while an application is running. In the current version messages have to be defined beforehand.

5.4 Calendar application

The last implemented application is a 3-D audio calendar reminiscent, based on the idea presented by Walker et al. [40]. The user can record calendar entries and place them into directions corresponding to the time of the day: 9 A.M. at -90 degrees azimuth (left of the user), noon at 0 degrees, three o'clock at +90 degrees (right of the user) and so on. The calendar application was implemented in two versions: head-tracked and non-head-tracked. The head-tracked version spins the calendar entries around a reference point (which is specified relative to the transmitter of the head-tracking system) and the non-head-tracked version obviously has the user's head as the only reference point.

The calendar application seems to work quite well, even though it certainly would help to have an easier interface to it, some kind of external controller device with which the user could browse through the calendar entries.

6 OVERVIEW OF RELATED TECHNOLOGY

6.1 Headsets

In order to send an acoustic environment from one place to another and then make it audible there is a natural need for microphones to record the environment or the communication and transducers to make the sound audible. As the system is aimed to be worn all the time, the transducers have to be small and light. Nowadays some people are wearing headphones throughout the day to listen to music. Adding a microphone to headphones could then be a logical way to make a two-way transducer system to record and reproduce the sound.

For augmented reality purposes it would be ideal if the surrounding and the augmented acoustic environment could be totally controlled with the headset. To achieve this the headset should block the surrounding sounds as effectively as possible and simultaneously play the acoustic environment back to the user. With headsets there are practically two ways to attenuate surrounding sound: in-ear passive earplugs and active noise cancelling headphones. Traditional passive hearing protectors would be effective but rather unpractical for this matter. There are various application-dependent headset systems available in the market. Most models rely on passive background noise rejection caused by tight fit of the headset inside the ear canal - though working like a traditional earplug. To get more attenuation at low frequencies some models have added electronics to cancel out background noise. There is a small microphone in the earplug and the signal it detects is fed back to the earplug in inversed phase. In some more advanced models there are two microphones and a separate DSP unit for processing the signals. The microphone signal is not available as an output in these headset systems. On the other hand, there are systems where the microphone signal is available either for communication or recording use. Table 1 lists some of the headsets available in the market, that have some relevance to the system considered in this article.

6.2 Head-trackers and location devices

Many of the applications discussed in this paper are based on knowledge of the location or orientation of the user. In the following presentation we roughly follow the taxonomy introduced in a recent review article [41].

In WARA applications, tracking of user's location and orientation takes place at least on two levels. First, in some applications we may be interested in the *global* po-

sition of a user, where the system is aware of the *physical* coordinates, such as in using the Global Positioning System (GPS), or the emerging European Galileo positioning system. Global information is typically needed in applications in which augmented acoustic events are localized to known real world objects (such as the Central train station, the parking lot of movie theater, or the department copying machine). An overview of different techniques operating at global, metropolitan, building, or room level are given in [41]. Global position information is needed to launch an application related to some object. In this article we put more emphasis on the problem of *head-tracking*. Estimates on the short-range relative location and orientation of the user's head are needed for successful augmentation of the acoustic environment.

The three major techniques for location-sensing are *triangulation*, *proximity* and *scene analysis*. Triangulation can be done via lateration, which uses multiple distance measurements to know points or via angulation, which measures bearing angle relative to points with known separation. In proximity technique the distance to a known set of points is measured and in scene analysis a view from a particular vantage point is examined.

GPS The Global Positioning System (GPS) is the most widely publicized location-sensing system at the moment. GPS is a satellite-based navigation system made up of a network of 24 satellites. The system is governed by the U.S Department of Defense. GPS satellites circle the earth twice a day in a very precise orbit and transmit signal information to earth. GPS receivers take this information and use triangulation to calculate the user's exact location. Essentially, the GPS receiver compares the time a signal was transmitted by a satellite with the time it was received. The time difference tells the GPS receiver how far away the satellite is. With distance measurements from a few satellites, the receiver can determine the user's position. The accuracy of the system is around 1-5 meters. The system doesn't work reliably indoors. [42]

GALILEO The Transport Council of the EU started GALILEO project in 1999. The system is a satellite-based global positioning system and is inter-operable with other systems such as GPS and UMTS. The GALILEO will cover all European states. The services offered are divided in three distinct services: General-

Purpose, Commercial and Public-Utility Services. The General-Purpose Services will be available for everybody without any need for authorisation. The accuracy will be 5-30 meters depending on technology. The Commercial Services provide an accuracy of 1-10 meters and are available as added value services on payment of a fee. The Public-Utility Services will provide a highly secure and accurate services for safety-of-life and other critical applications. The accuracy will be 1-6 meters. By the end of 2007, GALILEO should be fully operational.[43]

Active Badges Active Badge is an indoor locating system developed at the AT&T Cambridge. The system consists of a central server, fixed infrared sensors around the building and badges worn by the users. The badges emit a globally unique identifier every 10 seconds or on demand. A central server collects this data and determines the absolute location of the badge by using a diffuse infrared technique. The system can only tell the room where the badge is. On the other hand, the badge may have globally unique ID (GUID) for recognition. Sunlight and fluorescent light interfere with infrared.[41]

Active Bats Also developed by the AT&T researchers, the Active Bat location system uses an ultrasound time-of-flight lateration technique to provide a physical location of the Active Bat tag. The tags emit ultrasonic pulses, synchronized by radio frequencies, to a grid of ceiling-mounted sensors. The distance measurement data is forwarded to a central unit that calculates the position. The accuracy is 9 cm at 95 percent of the measurements. The tags can have GUID's for recognition.[41]

Cricket Instead of central computing, the Cricket Location Support System lets the user calculate the location. The infrastructure consists of infrared transmitters and the user objects act as receivers. Radio frequencies are used for synchronization and to delineate the time region during which the sounds should be received by receiver. A randomized algorithm allows multiple uncoordinated beacons to coexist in the same space. Cricket implements both the lateration and proximity techniques. The user can be located in around an area of 1.2 x 1.2 square meters within a room. Because all the computation is done at user side the power issues in wearable devices may become a burden.[41]

RADAR RADAR, developed by the Microsoft Research group, is an RF-based location system using the IEEE 802.11 WaveLAN wireless networking technology. At the base station, the system measures the signal strength and signal-to-noise ratio of signals sent by wireless devices. This data is then used in computing the 2D position within the building. There are two implementations for RADAR: one uses scene analysis and another uses lateration. In scene analysis the accuracy is around 5 meters and in lateration 4.3 meters in 50 percent of the measurements. The system requires at least three base stations per floor. There are also some companies offering 3-D indoor position tracking. [41]

SpotON SpotOn is an example of an ad hoc application where no infrastructure is implemented. The system implements ad hoc lateration with low-cost tags. SpotON uses radio signal attenuation to estimate intertag distance. Repeating the measurement with a few tags provides more accuracy. The achieved accuracy depends on the amount of tags. Ad hoc systems could be made to communicate with other, fixed, location systems to yield more accurate and usable positioning. [41]

6.3 Head-trackers

Head-tracking technologies can be roughly divided into acoustic, electromagnetic, inertial, mechanical, and optical tracking. Any of these techniques can be used to make a 6-degrees-of-freedom tracker (x, y, and z for position and yaw, pitch, and roll for orientation).

Acoustic tracking uses stationary microphones, and one or more movable 'user' elements. The user's movable elements emit high-frequency sound that is detected by a collection of stationary microphones. By measuring the time for sound to travel from transmitter to receivers it is possible to calculate the distance. The orientation is achieved by using multiple sensors. Because it takes some time for sound to travel from the transmitter to the sensors, the time-of-flight trackers suffer from slow update rate. Also, the speed of sound is affected by the environment variables (temperature, pressure,...) and this may occasionally cause problems. Acoustic tracking can also be utilized by using so called phase coherence tracking. In phase coherence tracking the stationary sensors detect the phase difference between the signal sent by the user emitter and a stationary emitter. As long as the distances are shorter than the wavelengths used, the distance can be determined. By using multiple sensors the

orientation can also be calculated. The phase coherence periodically updates the position and over some time error may accumulate in the location.

In *electromagnetic* tracking, a stationary element emits a pulsed magnetic field. A movable sensor attached to the user senses the field and reports the position and orientation relative to the source. Electromagnetic tracking can be accurate to within 2.5 mm in position and 0.1 degree in rotation, although accuracy deteriorates with distance from the source. Electromagnetic tracking is also susceptible to interference from metallic objects in the environment.

Inertial tracking devices represent a different mechanical approach, relying on the principle of conservation of angular momentum. These trackers use a couple of miniature gyroscopes to measure orientation changes. If full 6-DOF tracking ability is required, an inertial tracker must be supplemented by some position tracking device. A gyroscope consists of a rapidly spinning wheel suspended in a housing. The mechanical laws cause the wheel to resist any change in orientation. This resistance can be measured and converted into the yaw, pitch, and roll values. Inertial trackers sometimes suffer from drifting from the exact location - an error that accumulates over time. Some kind of occasional position verification would fix this problem.

Mechanical tracker devices measure position and orientation by using a direct mechanical connection between a reference point and a target. This could be done for example by an arm that connects a control box to a headband. The sensors at the joints measure the change in position and orientation with respect to a reference point. Mechanical trackers in general are very fast and accurate.

Optical trackers come in two variants. In the first one, one or several cameras are mounted on top of the headset, and a set of infrared LEDs is placed above the head at fixed locations in the environment. In the alternative setup, the cameras are mounted on the ceilings or a fixed frame, and a few LEDs are placed at fixed and known positions on the headset. In both approaches, the projections of the LEDs on the cameras image planes contain enough information to uniquely identify the position and orientation of the head. Various photogrammetric methods exist for computing this transformation. Optical trackers in general have high update rates and sufficiently short lags. However, they suffer from the line of sight problem, in that any obstacle between sensor and source

seriously degrades the tracker's performance. Ambient light and infrared radiation also adversely affect optical tracker performance.

Table 2 lists some headtrackers available in the market.

6.4 Estimation of room acoustical parameters

Acoustic parameters, such as amount of reverberation and geometric properties of the real environment around the user, may be needed for successful augmentation of the pseudo-acoustic environment. There are basically two mechanisms for acquiring this information. In some applications this information can be derived off-line and it would be available for the system as information related to that particular location. This would rely on a global localization system and an extensive database. This is basically the approach taken in the LISTEN project [19]. A more generic alternative would be based on on-line estimation of acoustic parameters from microphone input signals. The latter approach may be more suitable for wearable devices.

7 CONCLUSIONS

In this article we presented an overview of the concept of wearable augmented reality audio (WARA). We defined the idea of WARA and compared it to previous works on virtual and augmented reality audio. Different types of applications were discussed and compared in terms of their signal processing and hardware requirements. We also presented an overview of related transducer technology and localization and headtracking devices available at the time of writing this article.

In the case of open ears a listener may perceive the acoustic environment in its natural form. When wearing a headset with binaural microphones and earphones the user is exposed to a modified representation of the acoustic environment around the user, which is called here *pseudo-acoustic* environment. A traditional example of a pseudo-acoustic environment is that of a person wearing a binaural hearing aid device. *Virtual* sound environment is a synthesized binaural representation or a binaural recording. In *augmentation* of user's acoustic environment virtual sound objects are rendered on the natural, or a pseudo-acoustic representation of the sound field around the user. In this article, we focused on the latter case where virtual sound objects are combined with a pseudo-acoustic representation of the acoustic environment around a user. This is obtained using specific headset where miniature microphone elements are integrated into small earphone elements. Pseudo-acoustic represen-

tation is produced by routing microphone signals to the earphones.

We argue that it is technically and perceptually easier to produce a convincing fusion of virtual sounds and the acoustic environment around a user if virtual sounds are embedded into a pseudo-acoustic representation than the natural acoustic representation with open ears. In preliminary listening test results presented in this article we found that for an idealized case of using in-situ HRIRs in producing virtual sound objects even experienced listeners typically cannot discriminate virtual sounds from test sounds coming from loudspeakers placed behind a curtain in a listening room. It even turned out that the difference in results is small between earplug type earphones which block the ear canal, and conventional in-the-ear earphones which attenuate direct sounds by only a few decibels. Users also reported relatively fast adaptation to the pseudo-acoustic representation. Results are encouraging and suggest that the proposed transducer system may provide a potential framework for the development of applications of wearable augmented reality audio.

Finally, we introduced a system developed by the authors for testing of different applications and techniques of WARA. It is based on a specific headset configuration and a real-time software system running on a Linux machine.

8 ACKNOWLEDGEMENTS

A. Härmä's work was partially supported by the Graduate School GETA and the Academy of Finland.

9 REFERENCES

- [1] R. Held, "Shifts in binaural localization after prolonged exposures to atypical combinations of stimuli," *J. Am. Psychol.*, vol. 68, 1955.
- [2] A. M. Turing, "Computing machinery and intelligence," *Quarterly Rev. Psychol. Phil.*, vol. 109, October 1950.
- [3] J. Blauert, *Spatial Hearing: The psychophysics of human sound localization*. Cambridge, MA, USA: The MIT Press, 1999.
- [4] G. Plenge, "On the differences between localization and lateralization," *J. Acoust. Soc. Am.*, vol. 56, pp. 944–951, September 1972.

- [5] B. M. Sayers and E. C. Cherry, "Mechanism of binaural fusion in the hearing of speech," *J. Acoust. Soc. Am.*, vol. 29, pp. 973–987, September 1957.
- [6] W. M. Hartmann and A. Wittenberg, "On the externalization of sound images," *J. Acoust. Soc. Am.*, vol. 99, pp. 3678–3688, June 1996.
- [7] N. Sakamoto, T. Gotoh, and Y. Kimbura, "On out-of-head localization in headphone listening," *J. Audio Eng. Soc.*, vol. 24, pp. 710–715, 1976.
- [8] T. Lokki and H. Järveläinen, "Subjective evaluation of auralization of physics-based room acoustics modeling," in *Proc. Int. Conf. Auditory Display*, (Espoo, Finland), pp. 26–31, July 2001.
- [9] B. G. Shinn-Cunningham, N. I. Durlach, and R. M. Held, "Adapting to supernormal auditory localization cues. II constraints on adaptation of mean response," *J. Acoust. Soc. Am.*, vol. 103, pp. 3667–3676, June 1998.
- [10] D. R. Begault, *3-D sound for virtual reality and multimedia*. New York, USA: Academic Press, 1994.
- [11] J. Huopaniemi, *Virtual Acoustics and 3-D Sound in Multimedia Signal Processing*. PhD thesis, Helsinki University of Technology, Laboratory of Acoustics and Audio Signal Processing, Espoo, Finland, 1999. Report no. 53.
- [12] L. Savioja, *Modeling techniques for virtual acoustics*. PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, Espoo, Finland, 1999. Report TML-A3.
- [13] T. Lokki, *Physically-based Auralization - Design, Implementation, and Evaluation*. PhD thesis, Helsinki University of Technology, Telecommunications Software and Multimedia Laboratory, report TML-A5, 2002. Available at <http://lib.hut.fi/Diss/2002/isbn9512261588/>.
- [14] H. Fletcher, "An acoustic illusion telephonically achieved," *Bell Laboratories Record*, vol. 11, pp. 286–289, June 1933.
- [15] M. Cohen, S. Aoki, and N. Koizumi, "Augmented audio reality: Telepresence/vr hybrid acoustic environments," in *Proc. of IEEE International Workshop on Robot and Human Communication*, pp. 361–364, IEEE, July 1993.
- [16] J. Donath, K. Karahalios, and F. Viégas, "Visiphone," in *International Conference on Audio Display April 2000*, ICAD, April 2000.
- [17] J. Virolainen, "Design and implementation of a stereo audio conferencing system," Master's thesis, Helsinki University of Technology, 2001.
- [18] M. W. Krueger, *Artificial Reality II*. Addison-Wesley, 1991.
- [19] G. Eckel, "Immersive audio-augmented environments: The listen project," in *Proc. of Fifth International Conference on Information Visualisation*, pp. 571–573, IEEE, 2001.
- [20] B. Bederson, "Audio augmented reality: A prototype automated tour guide," in *Human Computer in Computing Systems*, pp. 210–211, ACM, 1995.
- [21] R. D. Shilling, T. Letowski, and R. Storms, "Spatial auditory displays for use within attack rotary wing aircraft," in *International Conference on Auditory Display April 2000*, (Atlanta, Georgia, USA), ICAD, April 2000.
- [22] N. Belkhamza, A. Chekima, R. Nagarajan, F. Wong, and S. Yaacob, "A stereo auditory display for visually impaired," in *IEEE TENCON 2000*, pp. II-377–382, IEEE, 2000.
- [23] D. P. Inman, K. Loge, and A. Cram, "Teaching orientation and mobility skills to blind children using computer generated 3-d sound environments," in *International Conference on Auditory Display April 2000*, ICAD, April 2000.
- [24] M. W. Krueger and D. Gilden, "Knowwheretm: An audio/spatial interface for blind people," in *International Conference on Audio Display November 1997*, ICAD, November 1997.
- [25] K. Lyons, M. Gandy, and T. Starner, "Guided by voices: An audio augmented reality system," in *International Conference on Auditory Display April 2000*, (Atlanta, Georgia, USA), ICAD, April 2000.

- [26] T. M. Drewes, E. D. Mynatt, and M. Gandy, "Sleuth: An audio experience," in *International Conference on Auditory Display April 2000*, ICAD, April 2000.
- [27] A. Walker, S. A. Webster, D. McGookin, and A. Ng, "A diary in the sky: A spatial audio display for a mobile calendar," in *Proc. BCS IHM-HCI 2001*, (Lille, France), pp. 531–540, Springer, 2001.
- [28] N. Sawhney and C. Schmandt, "Design of spatialized audio in nomadic environments," in *International Conference on Auditory Display November 1997*, ICAD, November 1997.
- [29] N. Sawhney and C. Schmandt, "Nomadic radio: Speech and audio interaction for contextual messaging in nomadic environments," *ACM Transactions on Computer-Human Interaction*, vol. 7, pp. 353–383, September 2000.
- [30] E. C. Cherry, "Some experiments on the recognition of speech, with one and two ears," *J. Acoust. Soc. Am.*, vol. 25, pp. 975–979, September 1953.
- [31] B. MacIntyre and E. D. Mynatt, "Augmenting intelligent environments: Augmented reality as an interface to intelligent environments," in *AAAI 1998 Spring Symposium Series, Intelligent Environments Symposium*, (Stanford University, CA), AAAI, March 1998.
- [32] E. D. Mynatt, M. Back, R. Want, and R. Frederick, "Audio aura: Light-weight audio augmented reality," in *The Tenth ACM Symposium on User Interface Software and Technology 1997*, (Banff, Alberta, Canada), ACM, October 1997.
- [33] V. C. L. Salvador, R. Minghim, and M. L. Pacheco, "Sonification to support visualization tasks," in *International Symposium on Computer Graphics, Image Processing, and Vision, 1998*, pp. 150–157, SIBGRAPI, 1998.
- [34] S. Helle, G. Leplâtre, J. Marila, and P. Laine, "Menu sonification in a mobile phone - a prototype study," in *Proceedings of the 2001 International Conference on Auditory Display*, (Espoo, Finland), pp. 255–260, ICAD, July 2001.
- [35] M. Li, H. G. McAllister, N. D. Black, and T. A. D. Pérez, "Perceptual time-frequency subtraction algorithm for noise reduction in hearing aids," in *IEEE Transactions on Biomedical Engineering* 2001, pp. 979–988, 2001, September 2001.
- [36] P. Runkle, A. Yendiki, and G. H. Wakefield, "Active sensory tuning for immersive spatialized audio," in *International Conference on Auditory Display April 2000*, ICAD, April 2000.
- [37] T. Ilmonen, "Mustajuuri - An Application and Toolkit for Interactive Audio Processing," in *Proceedings of the 7th International Conference on Auditory Display*, 2001.
- [38] L. Savioja, J. Huopaniemi, T. Lokki, and R. Väänänen, "Creating interactive virtual acoustic environments," *Journal of the Audio Engineering Society*, vol. 47, pp. 675–705, Sept. 1999.
- [39] R. Väänänen, V. Välimäki, J. Huopaniemi, and M. Karjalainen, "Efficient and parametric reverberator for room acoustics modeling," in *Proc. Int. Computer Music Conf. (ICMC'97)*, (Thessaloniki, Greece), pp. 200–203, Sept. 1997.
- [40] A. Walker, S. Brewster, D. McGookin, and A. Ng, "Diary in the Sky: A Spatial Audio Display for a Mobile Calendar," in *Proceedings of BCS IHM-HCI 2001*, pp. 531–540, 2001.
- [41] J. Hightower and G. Borriello, "Location systems for ubiquitous computing," in *Trends, Technologies & Applications in Mobile Computing* (N. Dimitrova, ed.), special report 5, pp. 57–66, IEEE Computer Society, 2001.
- [42] "Gps primer." <http://www.aero.org/publications/GPSPRIMER/index.html>.
- [43] "Global satellite navigation services for europe," in *GALILEO Definition Summary*, esa, European Space Agency, 2000.
- [44] "Invisio and bluespoon communication earplugs." <http://www.nextlink.to>.
- [45] "KOSS in-ear headsets." <http://www.koss.com>.
- [46] "Etymotic headsets." <http://www.etymotic.com>.
- [47] "Okm microphones." <http://www.okmmicrophones.com>.
- [48] "Sintef." <http://www.sintef.no/units/informatics/RoSandSoS/>.

- [49] “Polhemus 3D tracking devices.” <http://www.polhemus.com>.
- [50] “Motion trackers for computer graphic applications.” <http://www.ascension-tech.com>.
- [51] “Intersense motion trackers.” <http://www.isense.com>.

Brand	Description
NEXTLINK.TO [44]	
Invisio	The Invisio is an in-ear communication earplug with an internal microphone. The microphone detects the jaw-bone vibration and this way is fairly insensitive to surrounding noise and other sounds as well. The plug is wire connected to radio unit.
Bluespoon	This is a wireless, bluetooth version of the Invisio. Available only for the right ear
Sony	
MDR-NC10	An in-ear noise cancelling headphone. The earplug fits tightly in ear canal and attenuates surrounding sound fairly well even without switching the noise-cancelling on. The plugs are a bit bigger than conventional earplugs. The noise cancelling is done by analog means. The noise-cancelling circuitry works fairly well at low frequencies.
MDR-NC11	This is a newer model of the NC-10. The earplug is much smaller and lighter. Otherwise similar to the NC-10.
MDR-EX70LP	This is a small in-ear earplug. The fits tightly inside the ear canal and attenuates surroundings fairly well, especially at higher frequencies. Similar to the KOSS's ThePlug.
KOSS [45]	
ThePlug	The plug consists of a foam-like cushion that fills the ear canal and this way blocks the surrounding sounds. Similar to the Sony's MDR-EX70LP.
Complug	A communication earplug with a microphone. The earplug is an in-ear type and background noise rejection is increased with an electric noise-cancelling circuitry
Etymotic Research [46]	
ER-4	A high-end in-ear headphone with a very high background noise attenuation (20-25 dB). The attenuation works passively. There slightly different models (binaural, power and stereo) for different applications.
ER-6	Similar to the ER-4 but with little less isolation from the background noise.
OKM [47]	
in-ear microphones	These are in-ear electret microphones for binaural recordings. There is no headphone in this system.
PARAT [48]	
Communication earplug	A communication earplug developed in the research institute SINTEF (Trondheim, Norway). The earplug consists of a miniature loudspeaker and two microphones. The other microphone is outside the plug and the other microphone is located inside the plug. The earplug has also a built-in microchip for signal processing and a built-in radio unit for wireless communication. In normal situations, he earplug passes the surrounding sounds through but when high noise levels occur, the earplug blocks the surrounding sounds. The users voice can be recorded via the outside microphone or when in noisy environment, via the inner microphone.

Table 1: Transducer review

Brand	Method	Range	DOF	Other
Logitech Head-Tracker	acoust.	1.6 m	6	max angle 100 degree cone
Polhemus [49] FASTRAK	magn.	1.2-2 m	6	max. 16 receivers
ISOTRAK	magn.	1.6-2.3 m	6	max two receivers
Ascension [50] 3D Bird	inertial	cable length	3	Only orientation, computing in a PC
Flock of Birds	magn.	3 m	6	Up to four sensor
Laser Bird	optical	2 m	6	High update rate (240 Hz), max angle ± 55 horizontal and ± 54 vertical to 1.2 m
pciBIRD	magn.	76 cm	6	Small sensors (8 and 22 mm), computing in a PC
miniBIRD	magn.	76 cm	6	Miniature sensors (10 x 5 x 5 mm)
Nest of Birds	magn.	91 cm	6	up to four sensors, USB
Motion Star	magn.	up to 3 m	6	Also a wireless version, primary for motion tracking up to five characters with 18 sensors
InterSense [51] 900-series	inertial	20 - 72 m ²	6	Also a wireless version available, up to four sensors simultaneously
600-series	inertial/ultrasonic	16 - 50 m ²	6	can be used wireless if only orientation is needed
300-series	inertial	cable length	3	Only orientation, up to four sensors
InterTrax	inertial	cable length	3	Only orientation, USB
InertiaCube	inertial	cable length	3	Only orientation, computing in a PC

Table 2: Headtrackers