Hindawi Publishing Corporation
Journal of Electrical and Computer Engineering
Volume 2012, Article ID 457374, 13 pages
doi:10.1155/2012/457374

Research Article

Digital Augmented Reality Audio Headset

Jussi Rämö and Vesa Välimäki

Department of Signal Processing and Acoustics, Aalto University School of Electrical Engineering, P.O. Box 13000, 00076 Aalto, Finland

Correspondence should be addressed to Jussi Rämö, jussi.ramo@aalto.fi

Received 4 May 2012; Revised 31 August 2012; Accepted 9 September 2012

Academic Editor: Athanasios Mouchtaris

Augmented reality audio (ARA) combines virtual sound sources with the real sonic environment of the user. An ARA system can be realized with a headset containing binaural microphones. Ideally, the ARA headset should be acoustically transparent, that is, it should not cause audible modification to the surrounding sound. A practical implementation of an ARA mixer requires a low-latency headphone reproduction system with additional equalization to compensate for the attenuation and the modified ear canal resonances caused by the headphones. This paper proposes digital IIR filters to realize the required equalization and evaluates a real-time prototype ARA system. Measurements show that the throughput latency of the digital prototype ARA system can be less than 1.4 ms, which is sufficiently small in practice. When the direct and processed sounds are combined in the ear, a comb filtering effect is brought about and appears as notches in the frequency response. The comb filter effect in speech and music signals was studied in a listening test and it was found to be inaudible when the attenuation is 20 dB. Insert ARA headphones have a sufficient attenuation at frequencies above about 1 kHz. The proposed digital ARA system enables several immersive audio applications, such as a virtual audio tourist guide and audio teleconferencing.

1. Introduction

The concept of augmented reality is defined as a real-time combination of real and virtual worlds [1]. Probably the most intuitive implementation of augmented reality is a visual see-through display, which shows the real world with extended virtual content. The same concept is used in augmented reality audio (ARA), which combines everyday sound surroundings and virtual sounds in real time. In order to do this, specially designed ARA hardware is needed. A static ARA environment can be implemented using a loudspeaker array, however, the real usefulness of ARA is achieved with mobility. A mobile version of ARA can be implemented either with a stereo headset where binaural microphones are integrated into the earphones or with bone conduction headphones, which leave the ear canals open [1, 2].

Only the ARA headset is in the scope of this paper. The objective of the ARA headset is that the microphones should relay the surrounding sounds unaltered to the earphones, that is, the headset should be acoustically transparent, with no difference to the real surrounding sounds when listened without headphones. The copy of the surrounding sounds that has gone through the ARA headset is called a *pseudo-acoustic* environment [1].

The binaural microphones of the headset should be placed as close to the ear canal entrance as possible in order to preserve the spatial information of the sound [3]. Thus, the headset was constructed from a pair of active noise canceling headphones, which consists of in-ear type headphones and integrated binaural microphones. Alternatively, it is possible to construct the ARA headset by installing a pair of small electret microphones on to the earphones. When the microphones are placed near the entrance of the ear canal, the modifications of the ambient sounds due to the user's upper torso, head, and outer ear are preserved quite accurately. However, the ARA headset itself creates alterations to the pseudoacoustic representation of the real environment, mainly due to the headphone acoustics and the change of the outer ear acoustics when the headphone is inserted into the ear canal. Thus, the headset needs additional equalization in order to provide acoustically transparent reproduction of the ambient sounds.

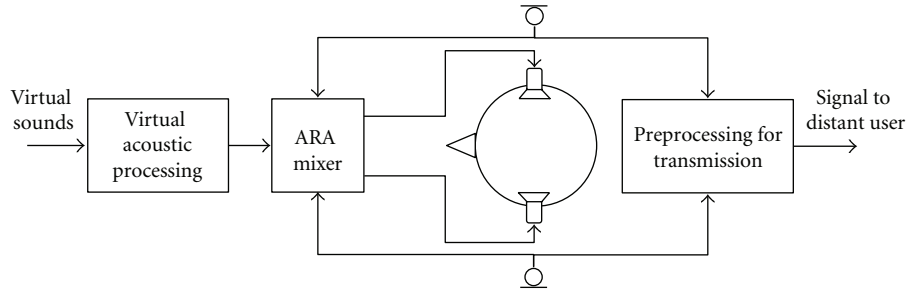

Figure 1: Block diagram of an ARA system.

Figure 1 shows the block diagram of a complete ARA system [1]. The first block on the left is used to create virtual sounds, which can be embedded into the pseudoacoustic representation. Furthermore, it can utilize location and orientation information to keep virtual audio objects in place while the user moves and turns his/her head. An essential part in creating a realistic ARA system is the ARA mixer, which routes and mixes all the signals involved in the system. Furthermore, the ARA mixer takes care of the equalization needed to create an acoustically transparent headset [4]. The headset is similar to common in-ear headphones, which are nowadays widely used with portable media players and smartphones. The preprocessing block can be used to send the user's binaural microphone signals to a distant user for communication purposes. The ARA mixer and headset are the user-worn devices of the system.

Even if the ARA headset would provide perfect sound quality and hear-through experience, it still requires useful applications in order to really benefit the user [5]. The most promising applications include full audio bandwidth (20 Hz–20 kHz) binaural telephony and audio conferencing with distant subjects panned around the user [6, 7]. Furthermore, the ARA technology enables location-based information services, such as virtual tourist guides and audio memos [1].

A previous prototype of the ARA mixer was constructed with analog electronics to avoid latency [8]. This is important because parts of the ambient sounds leak through and around the headset into the ear canal and if the pseudoacoustic representation is delayed by the ARA mixer, it results in a comb-filter effect when added to the leaked sound. However, there is a great interest in replacing the bulky and expensive analog components with digital signal processing (DSP). The digital implementation would bring several benefits when compared to the analog implementation. The benefits include programmability, which would enable a convenient use of individualized equalization curves; ease of design; precision. The downside is that a digital implementation introduces more delay than an analog implementation, which causes the comb filtering effect to the perceived signal. However, a digital implementation of the ARA mixer can be realized using a low-latency DSP due to the pronounced attenuation capability of the in-ear headset, which can dramatically reduce the comb-filter effect.

The aim of this paper is to study whether the ARA equalizer can be implemented using DSP and whether the latency between the pseudoacoustic representation and leakage of the headphone deteriorate the perceived sound excessively. The digital implementation of the ARA equalizer could bring many enhancements compared to the analog implementation, but only if the sound quality remains sufficiently good.

This paper is organized as follows. Section 2 describes the principles of the ARA technology. Section 3 concentrates on digital filters and their latency properties. Section 4 presents the group delay estimation of a passive mechanism. Section 5 focuses on the implementation of the digital ARA equalizer. Section 6 introduces a case study of a digital ARA mixer, and Section 7 concludes the paper.

2. ARA Technology

The ARA hardware has been specially designed and built for this purpose [4]. It consists of the ARA headset and the ARA mixer. The basis of the ARA headset is that it must be able to accurately reproduce the surrounding sound environment. In order to do that, the headset has two external microphones in addition to the earphone drivers. The quality of the reproduction of the pseudoacoustic environment must be sufficient enough for allowing the users to continuously wear the ARA headset for long periods of time nonstop.

However, because of the headphone acoustics, the pseudoacoustic representation is not an exact copy of the surrounding sounds. Thus, an equalizer is needed to correct the pseudoacoustic representation. Originally, the equalization was designed to be analog in order to have as low latency as possible [4]. Furthermore, the ARA mixer is used to embed virtual sound objects into the user's sound environment as well as to connect all the additional devices into the ARA system.

2.1. Headphone Acoustics. In normal listening with open ears, the incident sound waves are modified by the listener's body, head, and outer ear. When sounds are recorded with a conventional microphone situated away from the body and played back through headphones, the modifications caused by the body of the listener are lost. However, when microphones are placed binaurally near the ear canal entrances, the majority of these modifications are preserved.

The main difference when using in-ear headphones compared to the listening with open ears is that the headphones occlude the ear canals completely. An open ear canal acts as a quarter-wavelength resonator, that is, like a tube with one

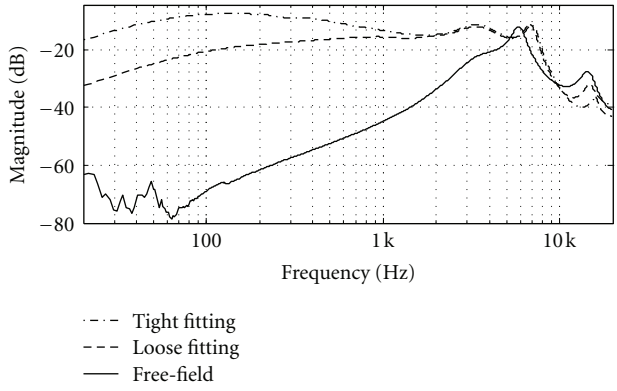

Figure 2: Frequency response of a headphone measured in free-field and in an ear canal simulator with tight and loose fitting (with 1/12 octave smoothing) [9].

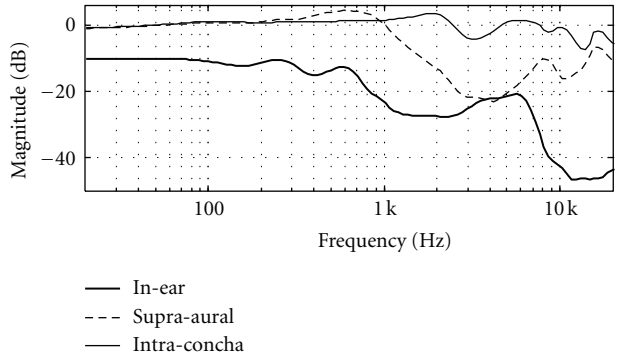

Figure 3: Isolation curves of different headphones (with 1/3 octave smoothing) [9].

end being open and the other end closed by the ear drum. The first quarter-wavelength resonance of the open ear canal occurs at approximately 2–4 kHz. When an in-ear headphone blocks the ear canal, it becomes a half-wavelength resonator. The closed ear canal does not only create a new half-wavelength resonance, but it also cancels the quarter-wave resonance created by the open ear canal, which people are accustomed to hearing. The first half-wavelength resonance occurs at approximately 5–10 kHz, depending on the length of the ear canal and the fitting of the headphone.

These resonance behaviors need to be taken into account when designing in-ear headphones. The basic idea in headphone design is to make the headphones sound natural, that is, as if one would listen with open ears. The headphones need to cancel the unnatural half-wavelength resonance and create the missing quarter-wavelength resonance in order to sound natural.

2.1.1. Leakage of the ARA Headset. Depending on the type of headphones, different amounts of ambient sound are transmitted to the ear canal as leakage around and through the headset. In this study we concentrate on the leakages of the in-ear type headphones used in the ARA headset. In some cases headphone leakage can be desirable, for example, when one needs to hear the surrounding environmental sounds. However, in the case of the ARA headset leakages, especially uncontrolled leakages, this can deteriorate the pseudoacoustic experience [10]. Leakages color the sound signal incident on the eardrum, since the sound signal that reaches the ear drum is the sum of the pseudoacoustic sound reproduced by the earphone transducer and the sounds leaked from the surrounding environment. The effects caused by leakage are most prominent at low and middle frequencies (below 1 kHz).

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

2.1.2. Pressure Chamber Principle. When listening with loudspeakers the sound pressure wave is radiated to the whole room around the listener. When using headphones, especially in-ear headphones, the volume of the space to which the wave is confined is extremely small compared to the volume of a room. The cavity between the in-ear headphone and the ear drum is about 1 cm^3. With such a small cavity it is easy to produce high sound pressure levels. At low and middle frequencies, where the wavelength is large compared to the earphone driver, the pressure inside the ear canal is in phase with the volume displacement of the transducer membrane and its amplitude is proportional to it [11]. Therefore, the pressure chamber principle enhances the low and middle frequencies.

In principle, to achieve the pressure chamber effect, the headphone should be tightly fit so that there would be no leaks, but in reality small leaks do not interfere much [11]. Figure 2 shows the frequency response of an in-ear headphone measured in free-field and in an ear canal simulator with tight fitting and with loose fitting. As can be seen, when the headphone is fit loosely into the ear canal simulator, the level of low frequencies decreases.

2.1.3. Headphone Attenuation. Figure 3 shows three different isolation curves measured from three different types of headphones. As can be seen, in-ear headphones are advantageous because they passively isolate the external ambient noise effectively. Furthermore, the isolation is highly frequency dependent. Additionally, a good fitting of the in-ear headphone is extremely important, because of the leakages that happen between the earphone cushion and skin deteriorate the passive isolation dramatically.

2.2. Equalizer. In order to find a proper equalization curve, the open ear case and the pseudoacoustic case were measured using the analog prototype of the ARA mixer [4]. Figure 4 shows the results of the measurement, where the black

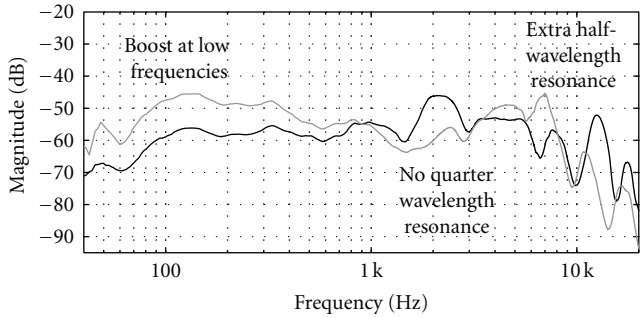

Figure 4: Transfer functions from an external sound source to an ear canal: black curve depicts an open ear case and gray curve is measured when the sound travels through an unequalized ARA headset into the ear canal. Modified from [4].

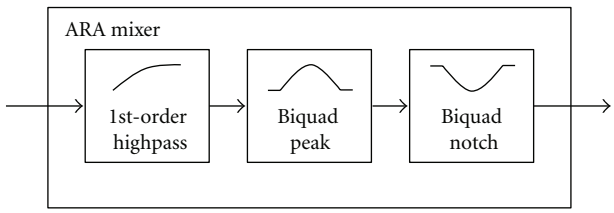

Figure 5: Equalization block diagram of the ARA mixer [4].

curve depicts the open ear case and the gray curve depicts the pseudoacoustic case, that is, the case where the sound has gone through the ARA headset and mixer. As can be seen, the pseudoacoustic representation has a boost at low frequencies (due to the pressure chamber principle), the quarter-wavelength resonance has disappeared, and the half-wavelength resonance has emerged (due to the closed ear canal). Hence, the equalization that is needed consists of a highpass filter which reduces the bass boost, a peak filter which restores the quarter-wavelength resonance, and notch filter which cancels the half-wavelength resonance. The block diagram of the equalizer is presented in Figure 5.

The current analog ARA equalizer uses a generic equalization curve that is the average of four person's individual equalization curves (see Figure 6). The highpass filter has an adjustable cutoff frequency from 6 to 720 Hz, the peak filter has a center frequency from 700 to 3200 Hz, and the notch filter has a center frequency that can be adjusted between 1800 and 8500 Hz [4].

2.3. Possible Applications. The ARA technology provides an implementation platform for innovative applications. Some of these application scenarios are briefly presented in this section. ARA applications can be categorized in many ways, including communication or information services, human-to-human, or machine-to-human communications [7]. For example, binaural telephony and teleconferences are human-to-human communication services, whereas a virtual audio tourist guide is a machine-to-human information service.

2.3.1. Binaural Telephony. Normal telephones and mobile phones transmit monosound and limit their bandwidth to

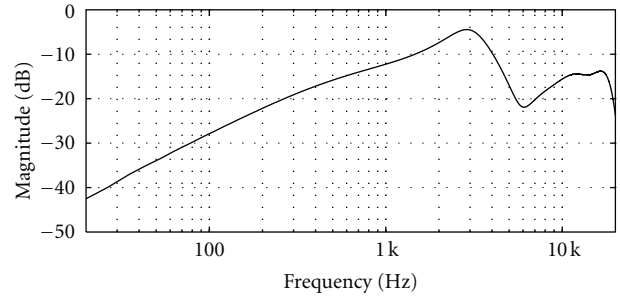

Figure 6: A generic equalization curve measured from the ARA mixer [4].

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

The inconvenience in the binaural telephony with the ARA headset is when either of the users talks, the voice is located inside the other user's head and in some situations the voice can be too loud as well. This can be avoided if the user's own voice is detected and then panned around the far-end user with the help of HTRFs. This creates a more natural feeling to the conversation because the far-end user appears in front of the listener [7].

2.3.2. Audio Teleconferences. Audio teleconferences are similar to binaural telephony except that there is usually a larger amount of participants than two. Nowadays teleconferences are common because of the globalized businesses, which often makes face-to-face meetings practically impossible. Traditionally audio teleconferences are held with the help of telephones and speakerphones. One of the problems here is the lack of telepresence because all the participants are reproduced through one phone or speakerphone. With the ARA headset audio teleconferences are brought to a new extent. It is easy to form discussion groups that consist of remote and local participants. Remote participants can be panned around the user (see Figure 7) and blended to the same acoustical environment as the user. This way it is much easier to separate the participants and distinguish who is talking.

There are many ways to utilize the ARA technology in different types of audio teleconferences. One scenario is that a traditional meeting is arranged and one team member is out of town but wants to participate. If he/she has an ARA headset and at least one person who is present at the meeting has an ARA headset, the out-of-town team member can participate in the meeting virtually (see Figure 8). Because

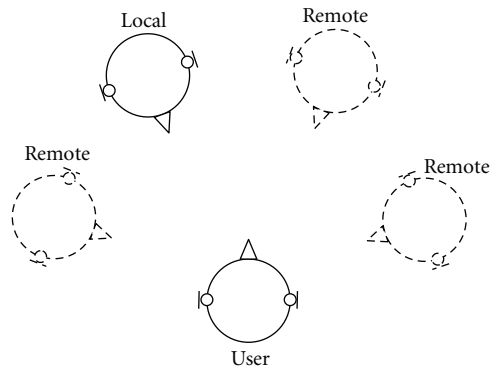

Figure 7: Diagram of audio teleconference using ARA technology where participants are panned around the user.

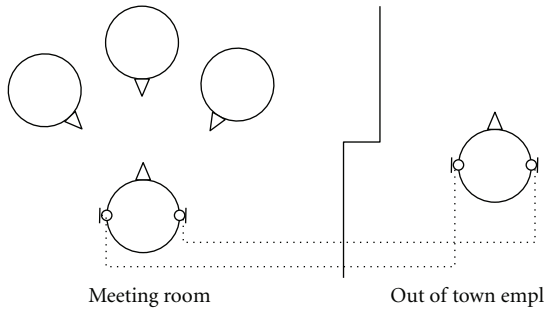

Figure 8: Diagram of audio teleconference where an out-of-town employee is participating the meeting with the help of ARA headset.

of the ARA technology, the out-of-town employee can hear exactly the same sound as the person who wears the ARA headset in the meeting room hears. One downside is that if there is no speaker system in the meeting room only the person who wears the ARA headset can hear the other team member.

Another inconvenience, similar to the binaural telephony, is that the near-end user's own voice gets localized inside the far end user's head and that the voice is much louder than the voices of the other participants in the meeting room. This can be avoided using the same principle as with the binaural telephony: voice activity detection and HRTF filtering [7].

2.3.3. Virtual Audio Tourist Guide. A virtual audio tourist guide is an interesting idea for an ARA application. It could replace tourist guides and give the user the freedom to explore a city by themselves without predetermined routes or schedules. The idea is that the ARA headset has a positioning capability, for example, GPS, and perhaps a head-tracker as well. Thus, the application knows where the user is and which way they are looking. The user can then walk around the city with the ARA headset and automatically hear information about interesting places they visit. Furthermore, the virtual audio tourist guide could have a GPS navigator application to guide the user from one place to another as well as contain all types of information about the city, such as restaurants, public transportation, and concerts [5].

3. Digital Filters

While analog filters use analog electronic components, such as resistors and capacitors, to perform the required filtering, digital filters use a computer or a signal processor to perform numeric calculations to sampled (discrete) signals. When digital filters are used, the analog sound signal must first be sampled and quantized, that is, transform the audio signal from sound to numbers. Sampling converts a continuous-time signal into a discrete-time signal and quantization maps the continuous magnitude values into discrete magnitude values. Sampling and quantization are performed using analog-to-digital converters (ADC). Furthermore, the filtered digital signal must be converted back to analog signal with a digital-to-analog converter (DAC) depicted in Figure 9.

Digital filters have many advantages compared to analog filters, such as the following.

(i) Digital filters are programmable, that is, digital filters can be easily changed without affecting the hardware.

(ii) With digital filters it is possible to create more precise and strict filters than with analog technology.

(iii) With digital filters it is straightforward to implement adaptive filters [12].

However, digital filters create much more latency than analog filters. In fact, analog filters have almost inconsequential latency, that is, the time that elapses when the electrical signal propagates through the filter circuitry.

3.1. Latency. Latency is defined as the elapsed time between a stimulus and the response [13]. With digital filters that means the time between the input and the output. The main sources of latencies in digital systems are filters, and the AD and DA converters (see Figure 9). Filters usually have frequency dependent delays, which can be illustrated with a group delay.

3.1.1. AD/DA Converters. The conversion of a continuous signal (voltage) into a sequence of numbers (digital signal) is called analog-to-digital (AD) conversion. The reverse process, that is, when a digital signal is converted back to the analog realm, is called digital-to-analog (DA) conversion. An AD converter first quantizes the analog signal in time, and then in voltage. It is important to do the quantizations in this order or otherwise it can cause gross errors [14]. The resolution of a converter implies the number of discrete amplitude values the AD converter can produce and it is usually expressed in bits. Other important properties of AD and DA converters are speed and accuracy.

AD converters have two types of latency: cycle latency and latency time. Cycle latency is defined to be the number of complete data cycles between the initiation of the input signal conversion and the availability of the corresponding output data. The latency time is the latency between the time where the signal acquisition begins to the time that the fully settled data is available to be read from the converter [15].

The conversion speed (in other words, latency) varies by the type of AD converter. Flash AD converters are considered

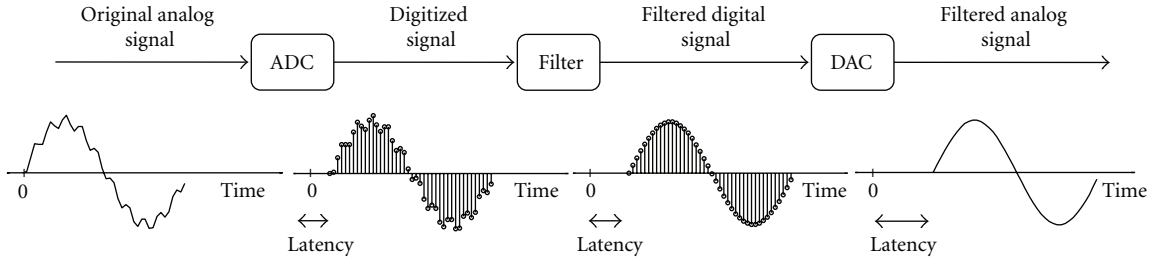

Figure 9: Basic functionality and latency of a digital filter.

to be the fastest type of AD converters. The basic idea of a flash AD converter is to compare the input voltage against a set of reference voltages, where the nearest value is selected to be the sampled value. In flash converters the conversion takes place in a single cycle [16]. However, flash AD converters are relatively expensive and their typical resolution is around 8–12 bits. The resolution is low because every one bit increase in resolution doubles the amount of the circuitry. However, straightforward increase in resolution can be obtained by stacking two flash converters. The delays in flash AD and DA converters can be, for example, 10 μs and 5 μs, respectively.

Common sound cards, such as an Edirol FA-101 FireWire audio interface or an internal sound card of a laptop, do not usually utilize very fast AD or DA converters. There is no particular need for faster converters, because they are mostly used for music listening, where a few millisecond delay is not a problem. For example, a simple measurement shows that the Edirol audio interface has a total delay of 21 ms and the internal sound card of a MacBook laptop has a delay of 8.6 ms. The measured delays include the delay of the DA converters (in the outputs) and the delay of the AD converters (in the inputs).

3.1.2. Phase and Group Delay of a Digital Filter. The real-valued phase response $\Theta(\omega)$ of a filter (i.e., the angle of the frequency response) gives the phase shift in radians that each input sinusoid component will undergo [17]. Two alternative delay responses can be derived from the phase response, namely, the phase delay and the group delay. While the phase response gives the phase shift in radians, the phase delay illustrates the time delay of each input sinusoidal component in seconds. The phase delay is defined as follows:

$$P(\omega) = \frac{\Theta(\omega)}{\omega}. \quad (1)$$

Whereas the phase delay gives the time delay of each sinusoidal component, the group delay gives the time delay of the amplitude envelope of a sinusoid, that is, the delay of a narrow band group of sinusoidal components. The group delay is defined as follows:

$$D(\omega) = -\frac{d}{d\omega}\Theta(\omega). \quad (2)$$

For linear phase responses the phase delay and group delay are identical, $P(\omega) = D(\omega)$ [17]. However, if the phase response is nonlinear, the phase and group delays differ. Furthermore, the phase delay can be considered to be the negative slope of a straight line fitted through 0 and a desired point ω of the phase response and the group delay can be considered to be the negative local gradient of the phase response.

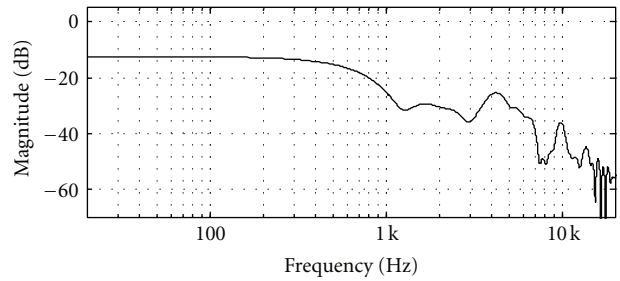

Figure 10: Isolation curve of an ARA headset.

4. Group Delay of a Passive Mechanism

Ambient noise attenuation is achieved when the in-ear headphone occludes the ear canal. The passive attenuation of a headphone behaves as a lowpass filter [18]. Figure 10 shows the isolation curve of an ARA headset measured with a head and torso simulator. The figure is showing an approximately 13-dB attenuation below the mechanical cutoff frequency and more than 20 dB of attenuation at higher frequencies.

As mentioned before the perceived sound in an augmented reality audio system is the sum of the pseudoacoustic representation of the surrounding sounds and the sounds that have leaked through the headset. The leaked sound has to propagate through the passive mechanism, that is, mechanical lowpass filter, and while doing that it is subject to an additional group delay of about 0.8–2 ms at below the mechanical cutoff frequency [18]. The mechanical group delay could allow additional electronic delay in the system, caused by the digital filters and AD/DA converters, without reduction in the performance of the ARA mixer.

A measurement was conducted to evaluate the delay caused by the passive sound transmission using two microphones. The first microphone was placed just outside the headset, while the second microphone was placed inside the ear canal. A sine sweep was played through an external sound source and the impulse responses of both microphones were measured. The cross-correlation between the outer and inner

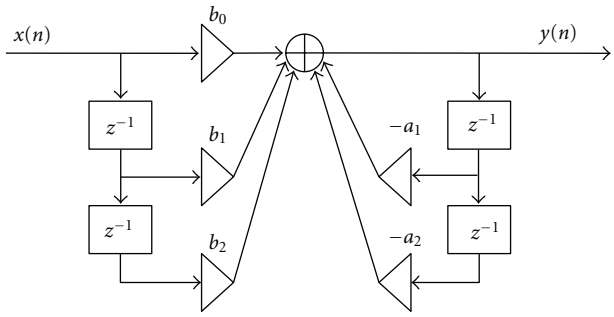

Figure 11: Direct form I implementation of a biquad IIR filter.

microphone was calculated while the headset was fitted first tightly and then very loosely into the ear. The obtained delays were approximately 0.21 ms for tightly fitted case and 0.06 ms for loosely fitted headset, which do not isolate the surrounding sounds at all. Thus, the delay is increased approximately 0.15 ms due to the mechanical response of the headset.

5. Digital Filter Design

There are two main types of digital filters, namely, FIR (finite impulse response) and IIR (infinite impulse response) filters. A FIR filter has a feedforward structure and it produces a finite impulse response. In contrast to FIR filters, an IIR filter has a feedback structure, which results in an infinite impulse response. FIR filters are inherently stable, whereas, IIR filters are not. The downside of FIR filters is that they require much more computation power compared to IIR filters. Furthermore, FIR filters introduce more latency to the signal because they usually have longer delay lines.

A biquad filter is an IIR filter type which can be used to create many varieties of filter response. The transfer function of a biquad filter consists of two quadratic functions:

$$H(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2}}{1 + a_1 z^{-1} + a_2 z^{-2}}, \tag{3}$$

where coefficients a_i and b_i determine the response of the filter. Figure 11 shows the direct form I implementation of a biquad filter. The direct form I structure is a straightforward approach for implementing an IIR filter. However, it requires $2N$ unit delays, when the filter order is N. There are also other possibilities to implement biquad filters, for example, the transposed direct form II, which requires only N unit delays for an Nth-order filter.

5.1. Digital Equalizer Design. One option is to use the measured impulse response of the analog ARA equalizer and create an FIR filter simply by using the values of the impulse response as filter coefficients. This technique yields a 45th-order FIR filter which has the same magnitude response as the original analog equalizer.

However, a more elegant way to implement the digital ARA equalizer is to design three independent IIR filters, that is, one first-order highpass filter (for bass control), one biquad peak filter (for creating the quarter-wavelength resonance), and one biquad notch filter (for canceling the half-wavelength resonance) as illustrated in Figure 5. The parametric IIR filter structure offers excellent adjustability, which enables useful features, such as individual equalization for different users.

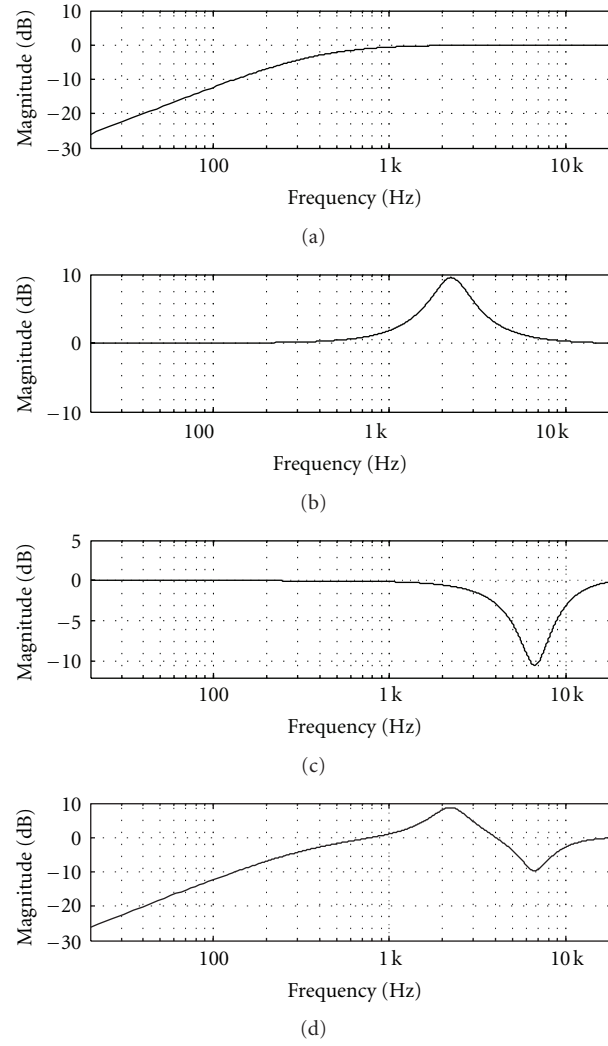

Figure 12: Magnitude frequency responses of the digital filters in the ARA equalizer: (a) the first-order Butterworth highpass filter, (b) the biquad peak filter, (c) the biquad notch filter, and (d) the combined response of the three filters.

Figure 12 shows the frequency responses of the three parametric IIR filters, as well as the combined frequency response of these filters, designed for the digital implementation of the ARA equalizer based on the analog ARA mixer. The topmost subfigure illustrates a first-order Butterworth highpass filter with a cutoff frequency of 400 Hz. The upper middle subfigure illustrates the biquad peak filter with a center frequency of 2250 Hz, a −3 dB bandwidth of 870 Hz, and a gain of 9.6 dB. The bottom middle subfigure shows the biquad notch filter while the center frequency is 6650 Hz, the −3 dB bandwidth is 2300 Hz, and the cut is 10.5 dB. The bottom subfigure shows the frequency response of the digital ARA equalizer, that is, the combined response of the three previously mentioned filters.

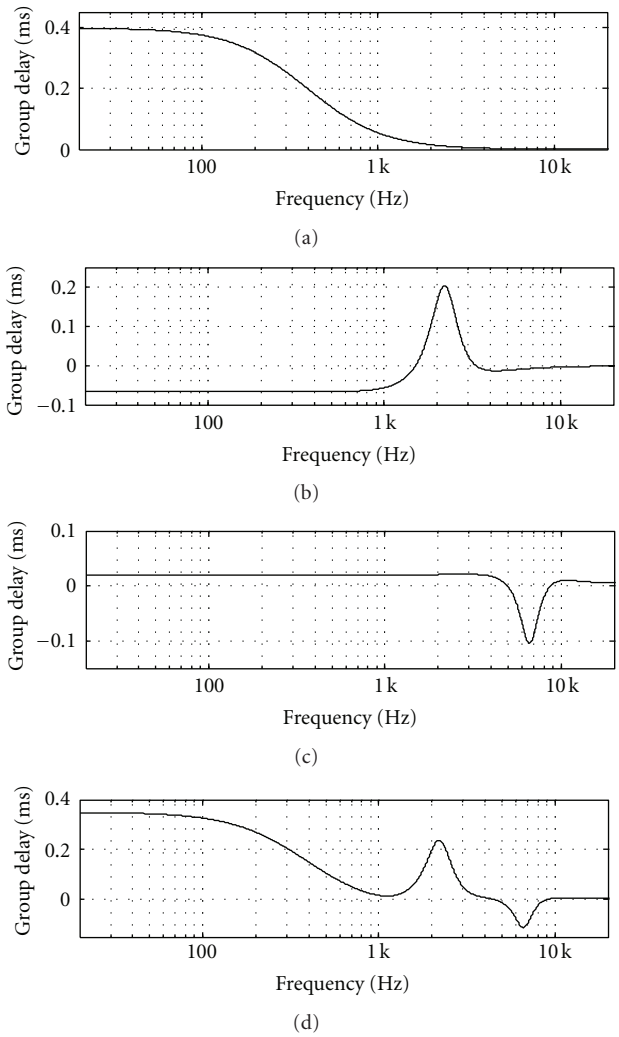

Figure 13: Group delays of the digital filters in the ARA equalizer: (a) the first-order Butterworth highpass filter, (b) the biquad peak filter, (c) the biquad notch filter, and (d) the combined delay of the three filters.

Figure 13 shows the group delays of the filters presented in Figure 12. The topmost subfigure shows the group delay of the first-order Butterworth filter, the upper middle subfigure shows the group delay of the peak filter, and the bottom middle subfigure shows the group delay of the notch filter, and the bottom subfigure shows the combined group delay of the filters.

As can be seen in Figure 13, the highpass filter (topmost subfigure) dominates the group delay at low frequencies, while the notch filter brings about an additional delay of 0.02 ms. The group delay of the peak filter is slightly negative (−0.06 ms) at low frequencies. Thus, the total group delay at low frequencies (20–1000 Hz) is approximately 0.35–0.02 ms. Furthermore, at higher frequencies, the group delay of the highpass filter is near zero and the peak filter dominates. The peak filter has a group delay of approximately 0.24 ms around its center frequency.

Table 1: Filter coefficients.

	Highpass	Peak	Notch
b_0	0.9723	1.1398	0.7655
b_1	−0.9723	−1.7658	−0.7736
b_2	0	0.7204	0.5645
a_1	−0.9446	−1.7658	−0.7736
a_2	0	0.8602	0.3300

6. Case Study

A DSP evaluation board, type ADAU1761 by Analog Devices [19], provides a complete DSP development environment. It includes the development software and the board itself, which has proper inputs and outputs as well as all the other hardware needed for the implementation. The evaluation board is a low-power stereo audio codec with integrated digital audio processing that supports stereo 48 kHz record and playback. The stereo ADCs and DACs support sample rates from 8 kHz to 96 kHz. The evaluation board uses a SigmaDSP core that features 28-bit processing (56-bit double precision). The sample rate used in this study was 44100 Hz.

In this section, a prototype implementation of an ARA system based on the Analog Devices DSP evaluation board is described and measurements of its delay characteristics are presented. The comb filtering effect caused by the combination of the processed and leaked sound in the user's ear is analyzed. Furthermore, a listening test, which was conducted to study the audibility of the comb filtering effect, is described.

6.1. Implementation of the Augmented Reality Audio Equalizer. The DSP programming is done with the SigmaStudio graphical development tool. SigmaStudio has an extensive library of algorithms to perform audio processing, such as filtering, dynamic processing, and mixing. Hence, it provides an easy way to implement DSP code via a graphical user interface.

Digital versions of the equalizer filters, discussed in the previous section, were implemented with Matlab by matching the digital filters with the analog ones. The filter coefficients are presented in Table 1 according to the transfer function of biquad IIR filter shown in (3).

Figure 14 shows the SigmaStudio implementation of the digital ARA equalizer. Starting from the left, the first block is the input block. Then there are three general filter blocks, namely highpass, peak, and notch, respectively. The IIR filter coefficients, shown in Table 1, are set accordingly to each filter block. Furthermore, there is a volume control for fine tuning the equalizer and finally the left and right output.

Figure 15 shows the measured frequency response of the digital equalizer (left channel). The measurement was conducted by connecting the input and output of the DSP board into an Edirol FA-101 audio interface and by playing a sine sweep through the equalizer [20]. A comparison of Figure 15 with the bottom part of Figure 12 shows a close correspondence.

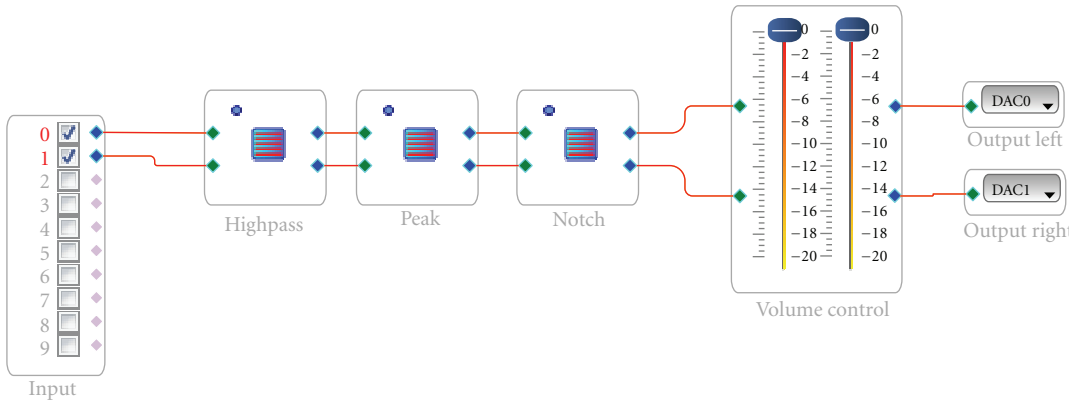

Figure 14: SigmaStudio implementation of the ARA equalizer.

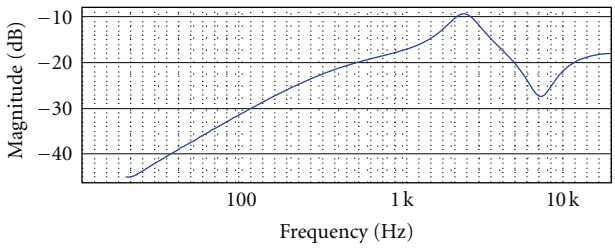

Figure 15: Measured frequency response of the digital ARA equalizer.

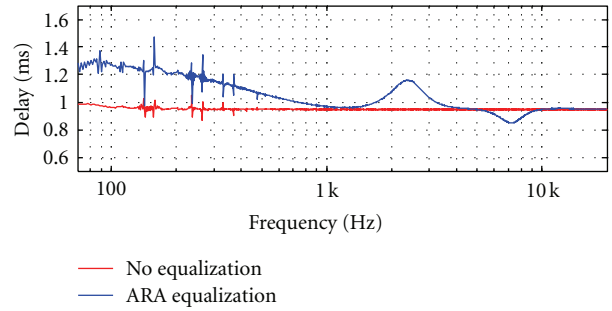

Figure 16: Group delay of the DSP Board with and without the ARA equalizer.

6.2. Measurements. A set of laboratory measurements were conducted in order to evaluate the suitability of the ADAU1761 DSP board as a digital ARA equalizer. The delay of the DSP board was measured in order to find out whether the DSP board has sufficiently fast AD and DA converters for a real-time application. Furthermore, the delay caused by the digital implementation of the equalizer was estimated.

6.2.1. Group Delay of the DSP Board. The measurement setup consisted of an MOTU UltraLite mk3 audio interface [21] and FuzzMeasure software [22]. A sine sweep was first played through the MOTU audio interface (the output was connected directly to the input) and after that through the MOTU audio interface and the DSP board. Furthermore, the DSP board had two configurations: one where the inputs were connected directly to the outputs, and the other where the implemented ARA equalizer was in use. Figure 16 shows the calculated group delays of these measurements, where the constant delay of the MOTU audio interface is removed. Thus, the measured group delays represent the delays that occur within the DSP board. The red curve depicts the case where the inputs of the DSP board is connected directly to the outputs, and the blue curve depicts the case where there was an implementation of the ARA equalizer between the inputs and outputs (see Figure 14).

As can be seen in Figure 16, without the equalizer (red curve) the group delay is approximately constant, about 0.95 ms. This result represents mainly the delay that is caused by the AD and DA converters. As a result 0.95 ms is quite good and this could allow the use of the ADAU1761 DSP board as a real-time ARA equalizer. As can be expected, the implementation of the equalizer slightly increases the group delay of the system. It is known that the group delay is increased at the stop band of a highpass filter (<1 kHz), as well as around the equalizer peak (1.5 kHz–4 kHz). The notch in the equalizer typically decreases the group delay, because the phase slope is positive around the center frequency of the notch (5 kHz–9 kHz). The measurement shows that the equalizer increases the group delay by less than 0.4 ms below 1 kHz and less than 0.2 ms above 1 kHz, as illustrated in Figure 16. The blue curve in Figure 16 is similar to the bottom part of Figure 13, as expected.

6.3. Comb Filtering Effect. When a delayed version of a signal is added to the signal itself, it causes a comb filtering effect. In the case of the ARA system, the latency of the pseudoacoustic representation should be very small, because the sound that has leaked through the ARA headset and the sound that has gone through the equalizer (pseudoacoustic representation) are summed at the ear drum. Thus, if the pseudoacoustic representation is delayed, it can cause a comb filtering effect.

The comb filtering effect in the ARA system is different than the usual loudspeaker and its reflection case, since the leaked sound and the pseudoacoustic signal can be controlled separately, that is, when the level of the pseudoacoustic signal is increased, the leaked sound remains the same. Furthermore, the objective is to reproduce the ambient sounds as

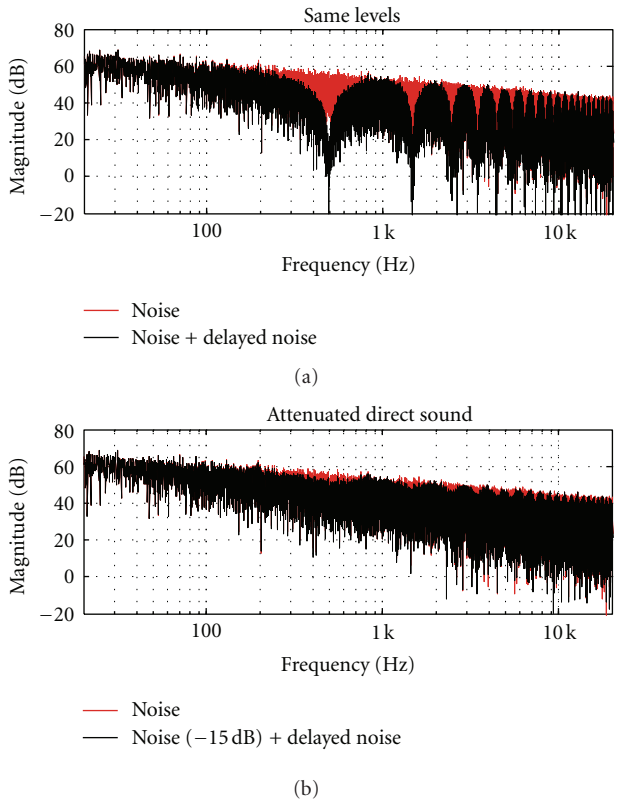

Figure 17: Comb filtering effect with the delay of 1 ms. The red curve represent the spectrum of pink noise and the black curve represents the spectrum of pink noise which is added to the delayed version of itself. (a) illustrates the comb filtering effect, when the added signals have the same energy, whereas (b) illustrates the effect when the passive attenuation of the ARA headset is considered.

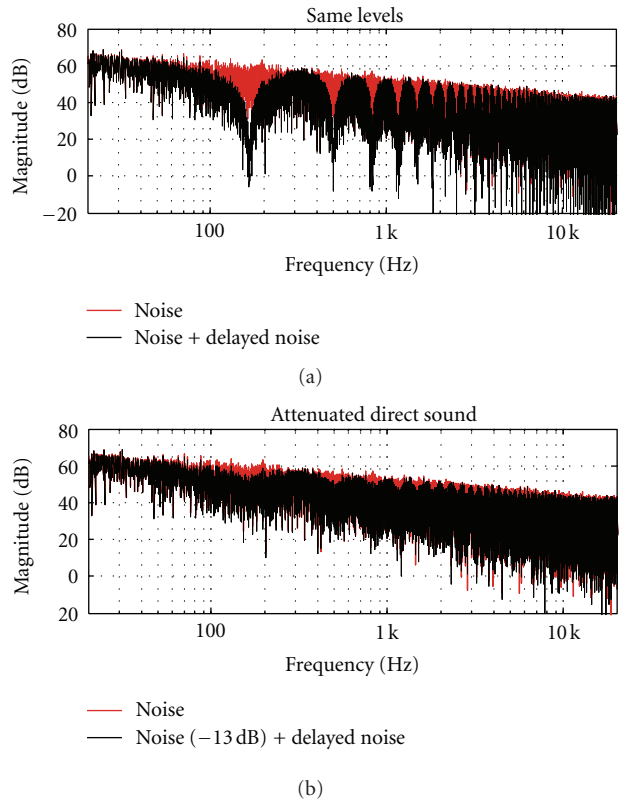

Figure 18: Comb filtering effect with the delay of 3 ms. The red curve represent the spectrum of pink noise and the black curve represents the spectrum of pink noise which is added to the delayed version of itself. (a) illustrates the comb filtering effect, when the added signals have the same energy, whereas (b) illustrates the effect when the passive attenuation of the ARA headset is considered.

unaltered as possible, which basically means that the combined level of the leaked sound and the pseudoacoustic sound should be the same as the ambient sound. The worst case scenario in terms of comb filtering effect occurs when the leaked sound and the pseudoacoustic sound have the same amount of energy. This happens when the headset attenuation is 6 dB.

The results from Figures 13 and 16 suggest that the group delay of the implemented digital ARA equalizer is below 1.4 ms. Figures 17 and 18 illustrate the theoretical comb filtering effects that could occur with the delays of 1 ms and 3 ms, respectively. The top subfigure in Figure 17 illustrates the spectrum of pink noise (red curve) and the spectrum of pink noise which is added to the delayed version of itself, with the delay of 1 ms (black curve). Furthermore, the signal levels are the same with both signals (noise and delayed noise). However, this is not the case in reality. The headset passively isolates the external sounds, thus, the nondelayed noise is quite heavily attenuated.

The first notch created by the comb filtering effect appears at the frequency of 490 Hz when the delay is 1 ms. Figure 10 shows that the passive attenuation of the ARA headset at 500 Hz is approximately 15 dB. Thus, the bottom subfigure in Figure 17 illustrates the same signals as the top subfigure, except that the nondelayed noise is attenuated 15 dB before the summation. The idea is that the nondelayed noise illustrates the leaked sound that has undergone the passive isolation of the headset and the delayed noise illustrates the pseudoacoustic representation delayed by the digital equalizer. Thus, the top subfigure represents the worst case scenario where the two signals have the same energy, that is, the headset attenuation is 6 dB and the bottom subfigure illustrates the case where the actual passive attenuation of the ARA headset is considered.

Figure 18 is similar to Figure 17, except that the delay is 3 ms instead of 1 ms. Thus, the first notch is created at the frequency of 170 Hz and the isolation of the headset is then 13 dB instead of 15 dB (see Figure 10). As can be seen from both of the figures, when the noise is attenuated according to the passive isolation of the ARA headset at the first notch frequency, the comb filtering effect diminishes dramatically (see the bottom subfigures). Furthermore, the passive isolation is even greater at frequencies above the first notch frequencies, thus, the effect of the comb filtering above these frequencies is actually even more faint than illustrated in the bottom subfigures.

Furthermore, Figures 19 and 20 show the frequency responses of FIR comb filters which correspond to the

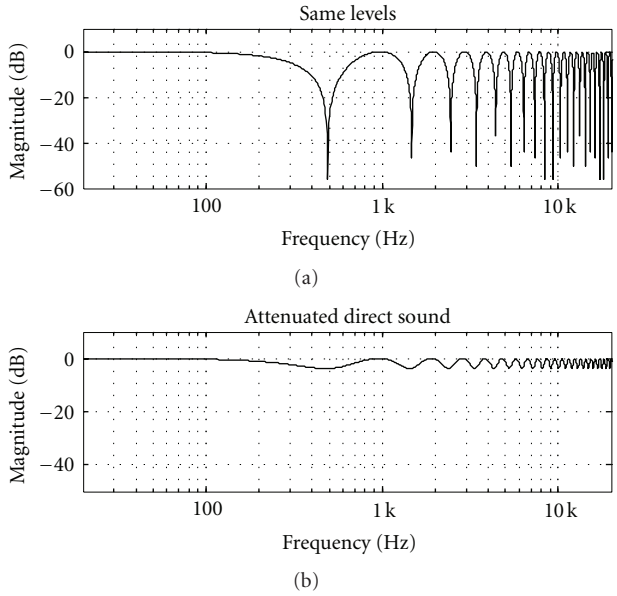

Figure 19: FIR comb filter, delay 1 ms. (a) illustrates the comb filtering effect, when the signals have the same energy, whereas (b) illustrates the effect when the passive attenuation of the headset (at the frequency of the first notch –15 dB at 490 Hz) is taken into account.

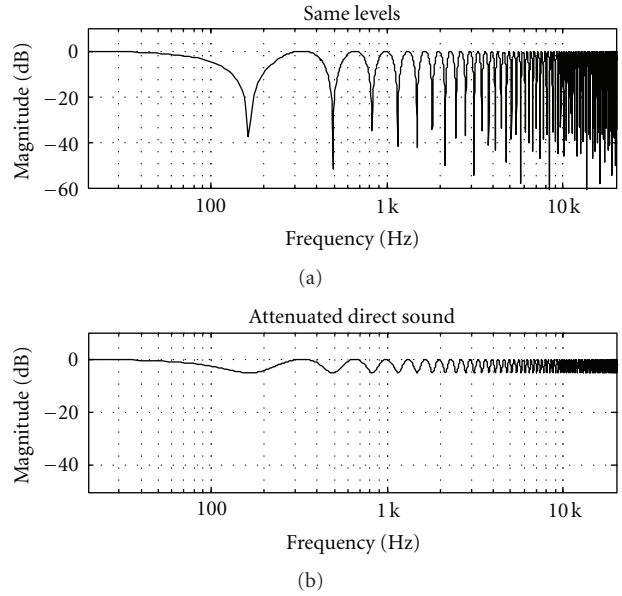

Figure 20: FIR comb filter, delay 3 ms. (a) illustrates the comb filtering effect, when the signals have the same energy, whereas (b) illustrates the effect when the passive attenuation of the headset (at the frequency of the first notch −13 dB at 170 Hz) is taken into account.

Figures 17 and 18. The transfer function of a FIR comb filter is

$$H(z) = g + (1 - g)z^{-L}, \tag{4}$$

where g is the gain for the direct sound (in this case, the leaked sound) and L is the delay in samples. For example, in Figure 19 the top figure has the values of $g = 1$ and $L = 44$, whereas the bottom figure has values of $g = 0.1778$ (-15 dB) and $L = 44$, with a sampling frequency of 44.1 kHz.

As can be seen from Figures 19 and 20, when the passive attenuation of the headset is taken into account, the comb filtering effect decreases dramatically. In fact, the deepest notches caused by the attenuated comb filtering effect is about 4 dB when the delay is 1 ms and about 5 dB when the delay is 3 ms. Note that the increase of the passive attenuation of the headset with the increasing frequency is not taken into account in these examples. Thus, the comb filtering effect actually decreases towards high frequencies.

6.4. Listening Test. The results of the previous section imply that the comb filtering effect should not be an insurmountable problem due to the attenuation of the ARA headset. It is known that narrow deep notches appearing in the frequency response are practically inaudible [23]. However, peaks in the frequency response lead to coloration. Thus, a formal listening test was conducted in order to subjectively evaluate the audibility of the comb filtering effect.

The listening test was conducted in the soundproof listening booth of the Aalto university. Two different test signals were used, namely, an English male speech and an instrumental music sample (beginning of Norah Jones—Don't Know Why). The length of the samples was approximately 4 seconds and they were played only once. The A-weighted sound pressure level (SPL) was 60 dB for the speech samples and 64 dB for the music samples. Seven subjects with normal hearing participated in the listening test. A pair of high-quality Sennheiser reference class HD 650 headphones was used. The test cases consisted of sample pairs, which included a reference sample (original sample) and a modified sample. The reference sample was always played back first and the question asked from the listeners was: "Is the sound quality of the second sample as good as that of the first sample?". The different test cases were presented in random order.

The modified samples were created by adding a delayed and attenuated version of the sample to the sample itself, causing the comb filtering effect. The used delays were 1, 3, and 5 ms, while the attenuations were 6, 14, and 20 dB, that is, nine different test cases per test signal. Furthermore, a reference sample pair was included in the test, where the listeners compared the reference signal to itself. All listeners evaluated each sample pair twice, except for the reference sample pair, which was included three times. The reference sample pair was included in order to be sure that the listeners actually could evaluate the samples correctly. It was decided that each listener should get at least two reference sample pairs correct out of the three in order to be considered a valid listener. All seven listeners fulfilled this criterion.

Figure 21 shows the results of the listening test, the black bars illustrate the speech samples and the white bars illustrate the music samples. As can be seen, the results for the speech and music signals are considerably different and the delay does not have as much effect as the attenuation does, which is consistent with the results of Brunner et al. [24]. Thus, the results can be divided into three groups based on

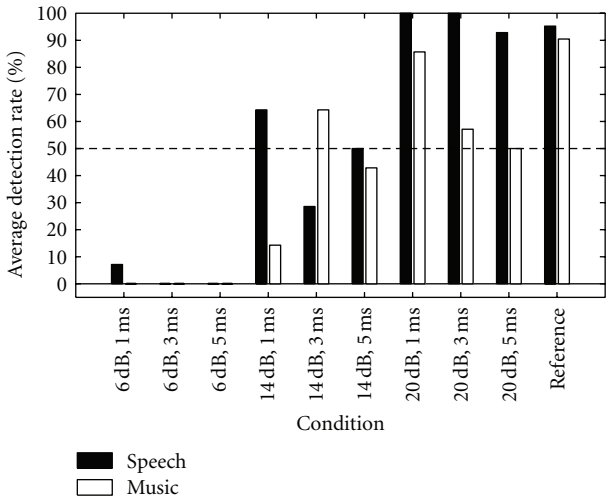

Figure 21: Results of the listening test.

the attenuation. When the attenuation was 20 dB, practically no one could hear any difference between the speech sample pairs and only a minority of the listeners heard a difference in music sample pairs. When the attenuation was decreased to 14 dB, listeners started to hear the timbre differences and when the attenuation was set to 6 dB, that is, the worst case scenario, all the listeners detected the degradation in the sound quality as expected.

However, in practical ARA situations such a straightforward comparison between the real ambient sounds and the pseudoacoustic representation is extremely rare. The idea in ARA is to wear the headset for long periods of time nonstop, which allows the user to adapt to the small timbre differences that the ARA headset introduces.

7. Conclusions

This paper has discussed and evaluated a digital implementation of a low-latency ARA mixer, which uses IIR equalizing filters. A headset with microphones in both ears is required. The ARA system combines virtual or transmitted sounds with the surrounding sounds, picked up by the ear microphones, and plays them to the user using the headphones. The system can be employed in various immersive augmented reality applications, such as in binaural telephony and audio teleconferencing.

It was argued in this paper that in-ear headphones are advantageous for implementing ARA systems, because they offer a high attenuation of sounds surrounding the user. Furthermore, the mechanical damping of the headphones is equivalent to a lowpass filter, which introduces an additional propagation delay. This extra delay is useful in ARA systems, because it allows latency in the processing of the microphone sounds before they are played to the ears of the user.

An ARA headset must include an equalizing filter, which compensates for the attenuation and changes in the ear canal acoustics caused by the headphones. Three cascaded IIR filters were proposed to implement the ARA equalization: a first-order highpass filter, a second-order peak filter, and a second-order notch filter. Filter coefficients for these filters were given. The overall throughput delay (latency) caused by the three filters is about 0.4 ms.

The ARA system was tested in real time using a DSP evaluation board. Measurements conducted with this prototype system show that the AD and DA converters and other electronics of the DSP board bring about a total delay of about 0.95 ms. When the proposed digital equalizing filter is inserted in the processing chain, the delay is maintained below 1.4 ms in the frequency range of 100 Hz–1 kHz and below 1.2 ms at frequencies above 1 kHz.

A listening test was conducted to learn about the audibility of the comb filtering effect, which is caused when the leakage of the direct sound through the headset is combined with the processed sound that is always delayed. It was found that the coloration caused by the comb filtering is inaudible in speech signals at all tested delays when the attenuation is 20 dB, but can be just audible in a music signal at some delay values. It is expected that the coloration is less disturbing in a practical case, when the user only hears the combination of the processed and leaked sound but a direct comparison against natural sound is unusual.

Future work includes the design of an equalizing filter to reduce the coloration caused by the combination of the leaked and processed sound at the user's ears.

Acknowledgment

The authors would like to thank Dr. Miikka Tikander for valuable comments and Mr. Julian Parker for his help in proofreading.

References

[1] A. Härmä, J. Jakka, M. Tikander et al., "Augmented reality audio for mobile and wearable appliances," *Journal of the Audio Engineering Society*, vol. 52, no. 6, pp. 618–639, 2004.

[2] R. W. Lindeman, H. Noma, and P. G. De Barros, "An empirical study of hear-through augmented reality: using bone conduction to deliver spatialized audio," in *Proceedings of IEEE Virtual Reality (VR '08)*, pp. 35–42, Reno, Nev, USA, March 2008.

[3] D. Brungart, A. Kordik, C. Eades, and B. Simpson, "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 (WASPAA '03)*, New Paltz, NY, USA, October 2003.

[4] V. Riikonen, M. Tikander, and M. Karjalainen, "An augmented reality audio mixer and equalizer," in *Proceedings of AES 124th Convention*, Amsterdam, The Netherlands, May 2008.

[5] M. Tikander, "Usability issues in listening to natural sounds with an augmented reality audio headset," *Journal of the Audio Engineering Society*, vol. 57, no. 6, pp. 430–441, 2009.

[6] A. Härmä, J. Jakka, M. Tikander et al., "Techniques and applications of wearable augmented reality audio," in *Proceedings of AES 114th Convention*, Amsterdam, The Netherlands, March 2003.

[7] T. Lokki, H. Nironen, S. Vesa, L. Savioja, A. Härmä, and M. Karjalainen, "Application scenarios of wearable and mobile

augmented reality audio," in *Proceedings of AES 116th Convention*, Berlin, Germany, May 2004.

[8] M. Tikander, M. Karjalainen, and V. Riikonen, "An augmented reality audio headset," in *Proceedings of the 11th International Conference on Digital Audio Effects (DAFx '08)*, Espoo, Finland, September 2008.

[9] J. Rämö and V. Välimäki, "Signal processing framework for virtual headphone listening tests in a noisy environment," in *Proceedings of AES 132nd Convention*, Budapest, Hungary, April 2012.

[10] M. Tikander, "Modeling the attenuation of a loosely-fit insert headphone for Augmented Reality Audio," in *Proceedings of AES 30th International Conference*, Saariselkä, Finland, March 2007.

[11] C. A. Poldy, "Headphones," in *Loudspeaker and Headphone Handbook*, J. Borwick, Ed., pp. 585–692, Focal Press, New York, NY, USA, 3rd edition, 1994.

[12] P. Titchener, "Adaptive filters for audio—an overview with application examples," in *Proceedings of the AES 93rd Convention*, San Francisco, Calif, USA, October 1992.

[13] D. Wessel and M. Wright, "Problems and prospects for intimate musical control of computers," *Computer Music Journal*, vol. 26, no. 3, pp. 11–22, 2002.

[14] M. Story, "Audio analog-to-digital converters," *Journal of the Audio Engineering Society*, vol. 52, no. 3, pp. 145–158, 2004.

[15] B. Baker, "Analogue-to-digital conversion: basics of ADC latency," *EE Times India*, 2009.

[16] M. Koen, "High speed data conversion," *Burr-Brown*, 1991.

[17] J. O. Smith, *Introduction to Digital Filters with Audio Applications*, W3K Publishing, 2007.

[18] B. Rafaely, "Active noise reducing headset—an overview," in *Proceedings of the International Congress and Exhibition on Noise Control Engineering*, The Hague, The Netherlands, August 2001.

[19] *ADAU1761 Data Sheet Rev C*, Analog Devices, 2010.

[20] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *Proceedings of AES 108th Convention*, Paris, France, February 2000.

[21] *UltraLite-mk3 Hybrid User Guide*, MOTU, 2010.

[22] *FuzzMeasure Pro 3.2 User Guide*, SuperMegaUltraGroovy, 2010.

[23] R. Buecklein, "The audibility of frequency response irregularities," *Journal of the Audio Engineering Society*, vol. 29, no. 3, pp. 126–131, 1981.

[24] S. Brunner, H.-J. Maempel, and S. Weinzierl, "On the audibility of comb-filter distortions," in *Proceedings of AES 122nd Convention*, Vienna, Austria, May 2007.