SOUND QUALITY OF AN AUGMENTED REALITY AUDIO HEADSET

Miikka Tikander

Helsinki University of Technology,
Laboratory of Acoustics and Audio Signal Processing,
Espoo, Finland
miikka.tikander@hut.fi

ABSTRACT

In augmented reality audio applications the user is exposed to a pseudo-acoustic reproduction of the real acoustic environment. This means that the surrounding sounds are heard through a specific augmented reality audio (ARA) headset [1]. Ideally the pseudo-acoustic environment should be an exact copy of the real acoustic environment.

The acceptability and usefulness of such a headset depends strongly on the overall sound quality of the headset. Especially if the headset is to be worn for longer periods of time in everyday life situations, the sound quality of the system must be sufficient.

In this paper an introduction to the aspects affecting the sound quality of an ARA-headset is addressed. In addition, results of a listening test for the overall sound quality of a well equalized headset are presented.

1. INTRODUCTION

The concept of *augmented reality audio* (ARA) headset has been introduced earlier e.g. in [1]. The basic idea in augmented reality audio is that a person's normal acoustic environment is enriched with virtual audio objects. This can be done by use of a special headset where binaural microphones have been integrated with headphones. When the microphone signals are directly routed to the earphones, the wearer is exposed to a binaural reproduction of the acoustic environment called the pseudo-acoustic environment. Ideally the user should not be able to hear any difference between the natural and the pseudo-acoustic environment.

This kind of a headset opens a possibility for a whole new set of application scenarios for mobile and wearable audio, see e.g. [2]. The headset provides a binaural audio I/O interface which can be exploited in many ways. The system could be worn for longer periods of time in everyday life situations, much like portable music players that are used daily by many users. However, in order to make this kind of a headset acceptable for daily usage, the sound quality of the pseudo-acoustic environment must be natural if not indistinguishable from the real acoustic environment.

This paper discusses some issues affecting the sound quality of ARA-headsets. Also the results from a listening test, where an overall sound quality of the pseudo-acoustic environment produced by a carefully equalized headset, are reported.

2. HEADSET

An ARA-headset has some basic requirements. First, the acoustic environment should be transmitted to the ear canal transparently. Secondly, the headset should provide accurate binaural signals from the user for further processing. Lastly, the headset should provide an interface to play sounds for the user.

2.1. Pseudo-acoustic environment

One strict requirement for the headset is that it should not alter the existing sound environment, thus the pseudo-acoustic environment should be a good enough copy of the existing sound environment. This way the user can interact normally with the natural surroundings. This sets high requirements for the binaural signals that are captured by the microphones and reproduced in the headset.

Depending on the type, the headset more or less blocks the ear canal and the recorded signals differ from signals that would exist at an unoccluded ear canal [3]. According to Algazi et.al. [4], HRTFs measured slightly outside a blocked ear canal still include all the spatial information that would exist at the ear drum in an unoccluded case. Thus, small earplug-type headsets would still provide all the necessary spatial information.

D'Angelo et al. studied how CIC (completely in the ear canal) hearing aids affect the localization ability [5]. A group of normal hearing people were wearing a CIC hearing aid that was equalized to have an identical transfer response compared to an unoccluded ear. According to results there was a small degradation in localization ability. Though, in the test the testee's head was fixed and there was no visual clues available. With ARA-headset the user has both visual and head movement clues together with the binaural signals and thus improving the spatial quality and accuracy of the pseudo-acoustics environment.

2.2. Earphone

The user is exposed to a real-time binaural recording from his or her own ears. The quality of the recording sets the lower limit to the available quality of the pseudo-acoustic environment. The next stage in the system is to play the binaural recording back to the user.

One common problem with headphones, especially with small earplugs that are fitted at the ear canal entrance or slightly inside the ear canal, is that the output properties depend on the fitting of the headset. The effect is strongest at low frequencies. This complicates any signal processing in the system.

Another phenomenon affecting the usability of an ARA-headset is the occlusion effect. When the ear canal is blocked the user's own voice tends to sound hollow or boomy. This is a common problem among hearing aid users. Although, this is an important usability issue, it will not be addressed any deeper in this report.

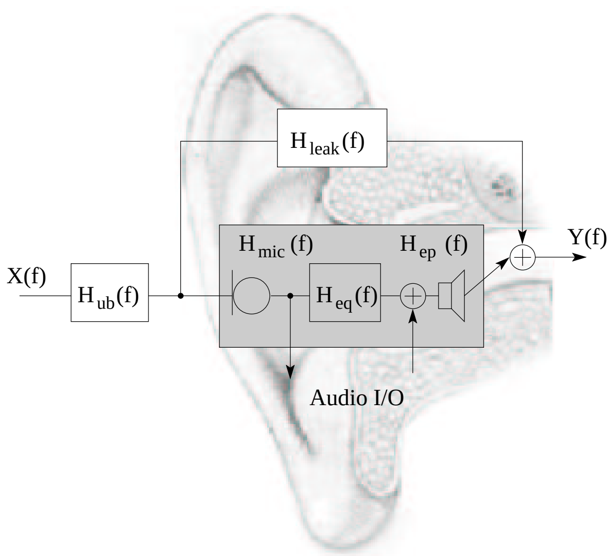

Figure 1: *Block diagram of an ARA-headset* .

2.3. Sound transmission through the headset

Fig. 1 shows a block diagram of the signal flow in an ARA-headset. The first block on the left $H_{ub}(f)$ accounts for the effect that the microphone placement and the geometry of the headset has on the sound field at microphone position.

The incoming sound signal goes through the headset but also leaks between the headset and the skin ($H_{leak}(f)$). The transmitted and the leaked sounds are summed in the ear canal. The leakage typically has a low-pass like characteristic and therefore the summed sound $Y(f)$ has low-pass like characteristics as well. $H_{mic}(f)$ and $H_{ep}(f)$ account for the microphone and the earphone transfer functions, respectively.

All of the above blocks colorate the transmitted sound to some degree, which has to be compensated somewhere in the system. The approach introduced here is to use an equalization filter $H_{eq}(f)$ to compensate any magnitude response errors in the system. If we assume that $H_{mic}(f) = H_{ep}(f) = 1$, the required filter $H_{eq}(f)$ can be written as:

$$H_{eq}(f) = \frac{1}{H_{ub}} - H_{leak}(f). \quad (1)$$

It should be noticed that the leakage path is delay-free and therefore it is essential that there should be no latency when processing $H_{eq}(f)$. In practice this means using analog filtering circuits, or very low latency DSP and AD/DA converters.

3. LISTENING TEST

A listening test was conducted to evaluate the overall sound quality of an equalized ARA-headset. The test was separated in two parts. First an equalization filter $H_{eq}(f)$ for the headset, used in the test, was measured. Then, with this filter, the overall sound quality of an ARA-headset was evaluated.

Fig. 2 shows the headset used in the listening test. The in-ear headset is a Voicetronic VT-01 (Fischer Amps) and the microphones are Senheisser KE-4 electret microphones. The reason for

Figure 2: *Headset used in the listening test.*

separate microphone and earphone unit is that originally the test was performed with two types of headsets (on-the-canal and in-ear), but the test data for the other headset was later found to be corrupted and thus results from one headset only is reported here.

3.1. Estimation of the equalization curve

The goal of this part of the test was to find the equalization filter $H_{eq}(f)$ for the headset used in the test. To get more information how required equalization differs between testees, a subjective equalization filter estimation method was chosen. An equalization curve was measured from each testee and the mean of these individual equalization curves was used as $H_{eq}(f)$ in the actual test. This non-individual filter was used to equalize the microphone signals prior feeding them to the headset, and thus (ideally) making the headset acoustically transparent.

3.1.1. Test procedure

The upper part of Fig. 3 (dashed lines) illustrates the setup for the equalization curve measurement test. The testee was wearing an ARA-headset in the right ear. Then a pair of circumaural headphones (Sennheiser HD600) was placed over the ears (including the ARA-headset).

With the circumaural headphones, a sequence of three identical sinusoids was played. The first and the last sinusoids were played for the left ear and the middle sinusoid was played for the right ear, which had the ARA-headset on (see Fig. 3). The testee's task was to set the sound level of the right channel to match the sound level on the left channel. By using the keyboard the testee was able to play the signals as many times as needed. The gain could be adjusted in 0.5, 1, and 3 dB steps. The frequency range covered thirty logarithmically spaced frequencies from 50 Hz to 18 kHz.

The test was first performed with both ears open (no ARA-headset). This was defined as reference measurement and was later used to compensate the actual measurement from any biasing due to the measurement setup. Then the measurement was repeated with the ARA-headset placed in the right ear.

Nine testees participated in the test. All of them work in the Laboratory of Acoustics and Audio Signal Processing at HUT. The testees had some experience on analytical listening. All of the testee's reported to have normal hearing. The fitting of the headset varied slightly between testees due to different sizes and shapes of

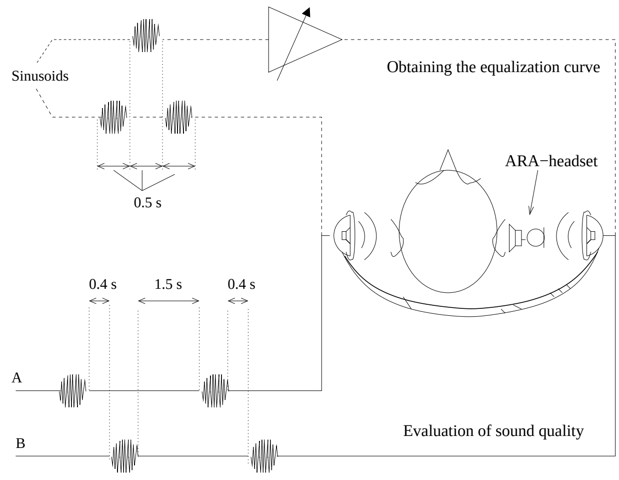

Figure 3: *Setup used in the listening test. The above signaling schematic was used in the equalization curve measurement. The lower part illustrates the actual listening test.*

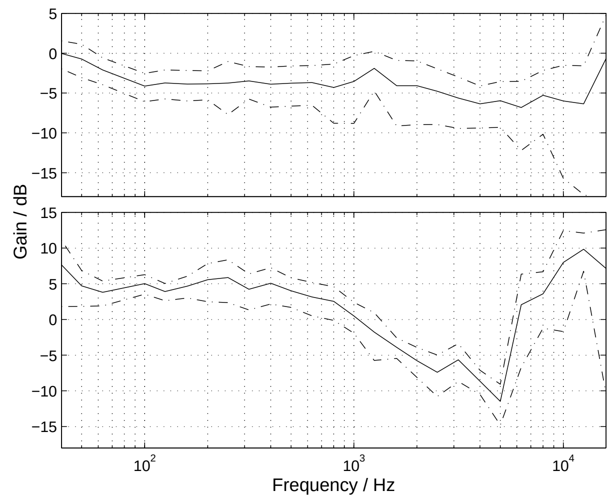

Figure 4: *Results (—) and 95 % confidence intervals (-.-) from the equalization filter estimation test. Upper panel: The reference case with no ARA-headset. Lower panel: Equalization curve for the VT-01 ARA-headset.*

the ears. For each testee the microphones were fitted as close to the ear canal entrance as possible (See Fig. 2).

3.1.2. The equalization filter

The upper panel in Fig. 4 shows the equalization filter estimation results in a case where the testees were wearing only the circumaural headphones. This measurement is used as a reference and to check the accuracy of the estimation method. Ideally the result should be a straight line but the curve seems to slightly decay towards higher frequencies. In the test procedure the sinusoids were played asymmetrically, twice for the left ear and once for the right ear. Loudness integration in hearing system requires more time at low frequencies and this could explain the emphasis at low frequencies in the results.

The results for the ARA-headset are shown in the lower panel of Fig. 4. The reference measurement data has been compensated out from the results. This equalization curve was used in the actual listening test. Only the magnitude spectrum was equalized.

3.2. Evaluation of sound quality

The goal of this part of the test was to evaluate the overall sound quality of pseudo-acoustic environment produced by an ARA-headset when the magnitude spectrum of the headset was equalized as described above. In other words, the goal was to determine how transparent the headset can be made by using a non-individualized equalizer to flatten the magnitude spectrum.

Eight persons participated in this part of the test. Seven of the participants belong to the group who took part in the first part as well.

3.2.1. Test procedure

The listening test setup is shown in the lower part of Fig. 3 (solid lines). The setup is identical to the one used in the first part. The testee was wearing an ARA-headset in one ear and the other ear was open. Then a pair of circumaural headphones was placed over the ears. Now, with one ear the testee was hearing a pseudo-acoustic reproduction of sounds played from the circumaural headphones, and the other ear was exposed to a natural acoustic reproduction of the circumaural headphones.

The test was run as an A/B paired comparison test where A was the reference and B was the pseudo-acoustic version of A (heard through the ARA-headset). The A/B sound sample pair was heard twice and then the testee had to evaluate the sound quality of B in comparison to A. The testee was instructed to consider the following attributes when judging the impairment: *timbre, clarity*, and *overall sound quality*. The testees were further instructed to consider the sound quality as if they had to wear the headset for longer periods of time in everyday life situations like going to music concerts, doing shopping, following lectures and so on. The only thing that was instructed not to take into account in judging was the inherent noise due the microphones in the headset. The impairment of B in reference to A was judged in a continuous scale from 1 - 5. The slider in the interface that was used for judging had numeric anchor points with written descriptions given by:

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

Values below four are considered as non-acceptable degradation of sound quality, and values from four to five are considered acceptable. The latter includes also the case where the test case (B) is found to sound better than the reference (A).

The equalization was done with an analog 31-band graphic equalizer (LA Audio EQ231G). The effective equalization response achieved with the equalizer was within $\pm$1.5 dB of the data shown in the lower panel of the Fig. 4.

3.2.2. Test cases and sound samples

Three different cases were tested:

Sample	Dur.	Abbr.
Shaking of a rattle in a steady rhythm	1.7	Ca
A hit on crash-like drum cymbal	1.6	Cy
An excerpt of rhythmic drumming	1.8	Bo
An excerpt of Danish male speech	2.7	M1
A phrase '*In language*' by an Am. male	1.0	M2
A sample of pink noise	1.1	PN
A pizzicato type trumpet playing	1.7	Tr
An accelerating bus by passing by	3.9	St

Table 1: Sound samples used in the listening test. The middle column lists the duration of a sample and the last column lists the abbreviations used in the figures.

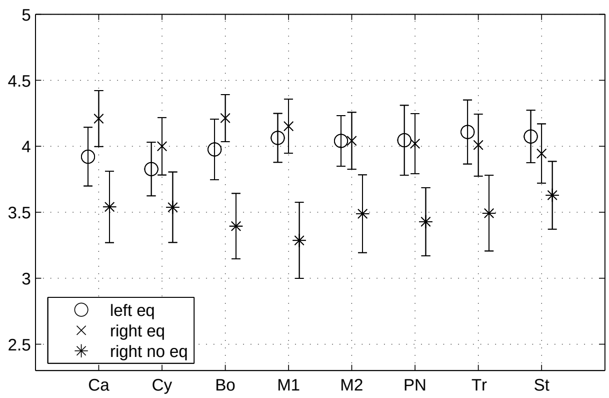

Figure 5: *Mean values of the grades for each sound sample with the ARA-headset and 95 % confidence intervals.*

- ARA-headset in the right ear and equalized
- ARA-headset in the right ear and unequalized
- ARA-headset in the left ear and equalized

One listening test session consisted of six trials, each having 24 sound samples to be graded. In one trial a set of eight sound samples were played three times in random order. One trial corresponded to one test case, so each test case was tested in two trials. The order of trials was randomized for every testee.

The sound samples used in the test are listed in Table 1. The samples consisted of natural sounds (musical instruments, speech, street noise) and also of one synthetic sound sample (pink noise).

4. RESULTS

Fig. 5 shows the results of the listening test. The data shown in the figure is the mean values and confidence intervals of the grades for each sound sample. Prior to computing the means the data was normalized by following the recommendation by ITU-R BS.1116-1.

All sound samples except St (street noise) seem to have benefited from magnitude equalization. However, confidence intervals are slightly overlapping with some samples. The St sound sample had most of its energy at frequency region where there was very little or no equalization at all, so this result was expected and thus gives more confidence to the results.

The equalization filter used in the test was an average of nine testees, measured from the right ear. However, using the same filter for each ear gave similar results. This is a good result, as the same filtering system can be used for both channels.

Another noteworthy result is that even the worst equalization result is still very close to grade four in MOS scale, which means that the overall sound quality is acceptable. However, the sound quality of the headset was fairly good to begin with and no conclusions can be drawn on how well this kind of equalization works with other types of headsets.

5. CONCLUSIONS AND DISCUSSION

It was found that with magnitude equalization the overall sound quality of an ARA-headset can be improved. In all cases (different sound samples) the sound quality was improved close to or above four in MOS-scale (perceptible, but not annoying degradation). This implies that this kind of a headset could be designed to have a sufficient sound quality to be used for augmented reality audio in everyday life situations, even for longer periods of time.

The equalization filter used in this test was an average measured from nine persons. There was a lot of variation in the individual results, and therefore this kind filtering works well for some subjects while for the others it may not work at all. For higher sound quality, one could use individual equalization curves. However, the acoustic properties of a headset are very sensitive to fitting of the earpiece. As the fitting changes also the required equalization filter needs to be changed.

One logical step in equalization would be to use adaptive equalization. This way any changes in acoustic properties of the headset due to changes between fittings could be automatically corrected in the equalization filter. One way to do this is to use two microphones, one outside the headset and one at the ear canal side of the headset. Naturally this requires more complex signal processing in the system.

6. ACKNOWLEDGMENTS

This work has been supported by Nokia Research Center and the Academy of Finland (project no. 53537).

7. REFERENCES

[1] A. Härmä, J. Jakka, M. Tikander, M. Karjalainen, T. Lokki, J. Hiipakka, and G. Lorho, "Augmented reality audio for mobile and wearable appliances," *Journal of the Audio Engineering Society (JAES)*, vol. 52, no. 6, pp. 618–639, June 2004.

[2] T. Lokki, H. Nironen, S. Vesa, L. Savioja, A. Härmä, and M. Karjalainen, "Application scenarios of wearable and mobile augmented reality audio," in *AES 116th Convention*. Berlin, Germany, May 8-11 2004.

[3] H. Møller, M. F. Sorensen, D. Hammershoi, and C. B. Jensen, "Head-related transfer functions of human subject," *J. Audio Eng. Soc.*, vol. 43, no. 5, pp. 300–321, May 1995.

[4] V. Ralp Algazi, Carlos Avendano, and Dennis Thompson, "Dependence of subject and measurement position in binaural signal acquisition," *J. Audio Eng. Soc.*, vol. 47, no. 11, pp. 937–947, November 1999.

[5] W. R. D'Angelo, R. S. Bolia, P. J. Mishler, and L. J. Morris, "Effects of CIC hearing aids on auditory localization by listeners with normal hearing," *Journal of Speech, Language, and Hearing Research*, vol. 44, pp. 1209–1214, December 2001.