Authenticity and Naturalness of Binaural Reproduction via Headphones regarding Different Equalization Methods

Josefa Oberem, Bruno Masiero, Janina Fels
Institute of Technical Acoustics, RWTH Aachen University, Germany, email: {job, bma, jfe}@akustik.rwth-aachen.de

Introduction

Not only a suitable localization performance, but also the plausibility and authenticity of the played scene are major criteria for a successful binaural reproduction. It is therefore important to examine whether the binaural reproduction can be perceptually distinguished from a real source. The aim of the presented investigation is to analyze the quality and reliability of binaural reproduction via headphones, especially for further investigations in a joint project on selective auditive attention [3]. Building on results of different measurements already published by Fels et al. [2], listening tests with a total of 80 participants were carried out. A similar approach with however a different focus has been published by Schärer and Lindau [4]. Moreover, two different microphone setups (miniature microphone in open dome and ear plug) used for individualized Head-Related Transfer Functions (HRTFs) and Headphone Transfer Function (HpTF) measurements were compared.

Method

To analyse the naturalness of the binaural reproduction via headphones two listening tests were designed. In the first one, the binaural reproduction was directly compared to the playback of real sources, while in the second one, an indirekt comparison was given.

Subjects A number of 80 students aged between 20 and 36 with normal-hearing participated voluntarily in the experiment. All listeners, 40 women and 40 men, can be considered as non-expert listeners.

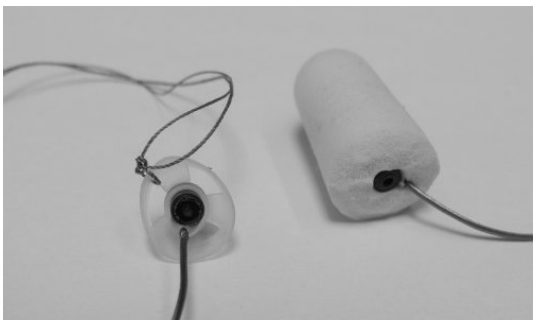

Figure 1: Open dome and ear plug with miniature microphone.

Microphones and Headphones Miniature microphones (*Sennheiser KE-3*) were fixed at the entrance of the ear with either an open dome (little silicon carrier) that does not change the impedance or an ear plug which closes the ear canal (cf. Fig. 1). To ensure a perfect fit of the microphone, the ear plug was shortened in length to be flush with the entrance of the ear canal. For the binaural reproduction open headphones (*Sennheiser HD 600*) were used.

Listening Test Setup The listening tests took place in a fully anechoic chamber ($l \times w \times h = 9.2 \times 6.2 \times 5.0$ m^3). The subject was asked to sit inside a frame of loudspeakers (cf. Fig. 2). A number of 24 loudspeakers were equally distributed over azimuth and three elevation levels, while the distance was kept constant at 1.7 m. The chair was provided with an adjustable head rest. To control the movements of the subject's head an electromagnetic tracker (Polhemus Patriot) was used.

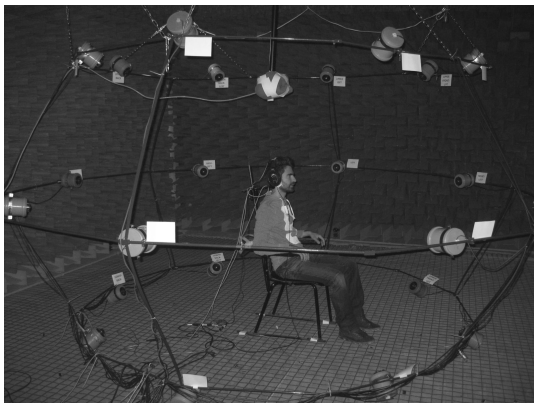

Figure 2: Construction of loudspeakers for listening test with participant sitting in chair with headrest.

Stimuli In listening test I three different stimuli were presented:

- Pulsed pink noise [200 Hz – 20 kHz] (0.8 s)
- Music [200 Hz – 10 kHz] (1.8 s)
- Speech [200 Hz – 8 kHz] (0.8 s)

On the other hand in listening test II the only stimulus used was the pulsed pink noise.

Binaural Recordings and Reproduction In this investigation HRTFs were measured individually for every subject and also statically from every loudspeaker of the described setup. In order to keep headphones on head during the presentation of stimuli in the listening test, subjects had to wear headphones also during the measurement. Based on the findings by Masiero and

Fels [1], but slightly changed a number of eight HpTFs were measured for every subject to calculate a robust headphone equalization.

Experimental Design

Subjects were split into two different groups. A number of 40 listeners belonged to the open meatus group and 40 listeners to the blocked meatus group.

Listening Test I: Direct Comparison Main task of this part of the test was to distinguish one stimulus out of three (3-AFC). The differences between the three choices was the reproduction method. For example two of the three stimuli building one set were played by loudspeakers, whereas the third wass binaurally reproduced by headphones. The aim was to check whether a difference can be heard. Therefore, no other changes between the stimuli were made.
The presented sets of stimuli were roved in level and equally distributed over all directions over all participants in 20 sets each. A participant's head movements were observed and in case of trespassing the given limits, sets were considered invalid and were repeated.

Listening Test II: Indirect Comparison In the second part of the listening test a pulsed pink noise was played, either by the loudspeaker or by the headphones. The subjects task was to decide whether the sound event was generated by the loudspeaker or the headphones (2-AFC). In listening test I participants were indirectly asked to indicate whether the loudspeakers or the headphones reproduced the stimulus, other than in this part, where subjects were directly confronted with the decision and did not have a reference.
Every subject listened to five stimuli played by headphones and five stimuli played by loudspeakers, which concluded in a total of ten sets.

Results

Listening Test I: Direct Comparison Figure 3 shows the results for all participants and all kinds of stimuli. Most of the subjects heard a difference between the pink noise delivered by real sources or as a binaural stimulus via headphones. In numbers, 18.49% (44 out of 238) of all sets of played noise stimuli with equalization in open ear canal were not answered correctly and 16.74% (38 out of 227) for the equalization in blocked ear canal. However, outliers show that there are participants that cannot distinguish between binaural reproduction and real sources. (While numbers given in the text are based on arithmetic mean, measures of central tendency for boxplots are medians.)
Since the listening test was a 3-AFC test, participants show a percentage of wrong answers of 66.67% when only guessing and therefore did not hear a difference. In case participants answered for one out of three times wrongly (33.33%), they could not hear a difference for 50% of the stimuli. In this case, in relation to all subjects, it can be said that 50% of all listeners did not hear any difference. The median for the categories of music lies exactly on the boarder of 33.33%. In calculations based on all errors without any respect to the deviation between participants the percentages are slightly higher: for the group of open meatus 34.76% (73 out of 210) and for the group of blocked meatus 35.10% (73 out of 208). Therefore at least 50% of all listeners could not distinguish between the binaural reproduction and the real sources.

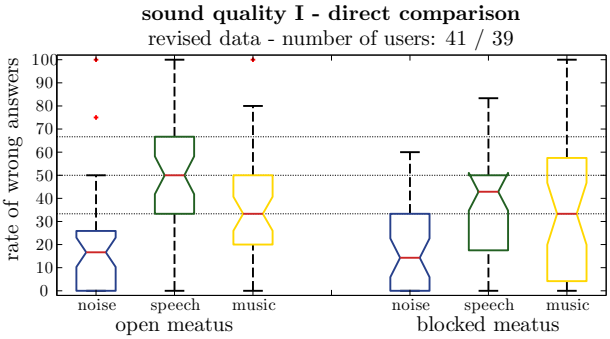

Figure 3: Box plots including median showing percentage of wrong answers for three stimuli (noise, speech and music) for two groups of equalization (open meatus, blocked meatus) of listening test I.

For the speech stimulus even more subjects were not able to hear any difference. While for the group of open meatus equalization nearly 75% of all participants chose the wrong answer, due to an error rate of 48.79% (121 out of 248), for the group of blocked meatus equalization it can at least be stated that 50% of the listeners could not perceive any difference (error rate: 38.17% (92 out of 241)).
On account of large variations within groups no significant difference can be seen between the two equalization methods regarding each of the three stimuli. However, the results for music and speech are significantly different than the ones for noise ($F = 10.77$, $p < 0.001$).

Listening Test II: Indirect Comparison For the listening test with an indirect comparison all participants were not able to distinguish between headphones and loudspeakers. Medians in Figure 4 lie at 50%. For calculations of arithmetic mean, results are similar and report 49.25% for the equalization in open meatus and 51.03% for the equalization in blocked ear canal. For a 2-AFC test a percentage of 50% shows that no subject was able to hear a difference. Variations are also smaller than in listening test part I and show just one single outlier, who answered falsely in all cases. The probability that he was able to hear a difference is large, but however he was not able to assign stimuli to the right reproduction method. Therefore, binaural reproduction sounded very real for all subjects.
Figure 5 shows that participants chose rather the loudspeaker (63.25%) as the reproducing method than the headphones (36.75%). While 32.00% of all stimuli presented by real sources were answered correctly, a rate of only 18.00% was observed for the binaural reproduction.

Since participants had two possible options to choose from, the rest to 50% was answered incorrectly. There is no difference between the two equalization methods and therefore the figure only shows the results for the open ear canal.

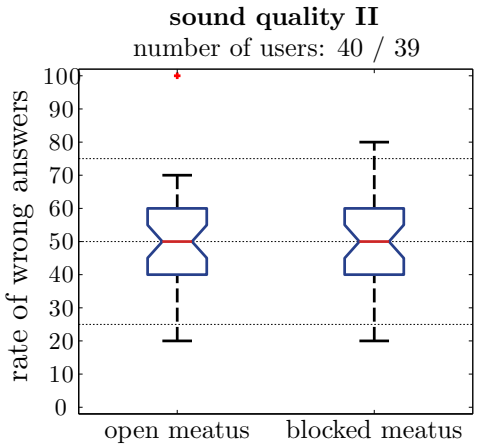

Figure 4: Box plot showing percentage of wrong answers of listening test II for equalization with open meatus and blocked meatus.

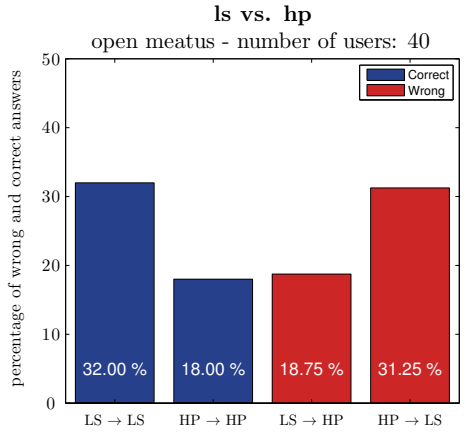

Figure 5: Histogram showing percentage of four combinations of playing reproduction methods and received reproduction methods.

Discussion

Listening Test I: Direct Comparison Results show that different numbers of subjects were able to distinguish between binaural reproduction and real sources according to the type of stimulus. Therefore the spectrum of the particular stimulus plays an important role. Numbers of errors show that stimuli with dominant high frequencies tend to be easier to distinguish than those without higher frequencies.
For pink noise, subjects reported that different colorations in higher frequencies were audible. Since frequencies higher than 10 kHz are very sensitive to the direction of incidence as well as to characteristics of headphones and therefore to the equalization, observations of subjects are reasonable.
Subjects mainly reported differences in direction of some degrees. Since possible rotations of head given by the head tracking system were greater than the localization blur, dislocations especially in front and back were very probable.
Summarizing, in a direct comparison between binaural reproduction and real sources half of the participants could not distinguish between the reproduction method for stimuli with dominant parts in frequencies below 10 kHz, like music or speech. Pink noise with a spectrum of 200 Hz - 20 kHz, coloration sounded different in the high frequency range for most of the subjects.
Finally, it needs to be mentioned that no significant difference between the equalization in open and closed meatus could be found.

Listening Test II: Indirect Comparison For listening test II, it can be concluded, that not a single subject out of 80 listeners was able to distinguish between real sources and binaural reproduction, when not being directly compared. As to see in the discussion of listening test part I, pink noise, due to its frequency spectrum, is the most sensitive stimulus used in this study to prove indistinguishability between binaural reproduction and real sources. Since subjects were not able to find differences for this stimulus, it can be concluded that subjects would also not able to distinguish between real sources and binaural reproduction for stimuli like music or speech.

Conclusion and Outlook

In future investigations [3] the binaural reproduction of speech is of major importance. Since results of listening test I show that 50% of all participants are not able to hear a difference, also confirmed by the results of the second part, the used binaural reproduction method including the equalization of headphones is applicable for further studies. The individual equalization can either be done with an open or a closed meatus.

References

[1] Masiero, B. and Fels, J.: Perceptually robust headphone equalization for binaural reproduction. Audio Engineering Society (2011), 8388:7

[2] Fels, J. and Vorländer, M. and Masiero, B. and Oberem, J. and Lawo, V. and Koch, I.: Experiments on cognitive performance using binaural stimuli. Inproceedings of Inter Noise, 2012

[3] Fels, J. and Oberem, J. and Karnbach, B. and Lawo, V. and Koch, I.: Comparison of dichotic and binaural reproduction in an experiment on auditory selective attention. Inproceedings of DAGA/AIA, 2013

[4] Schärer, Z. and Lindau, A.: Evaluation of equalization methods for binaural signals. Inproceedings of AES 126th Convention (2013), 7721