Applied Acoustics 114 (2016) 71–78

Contents lists available at ScienceDirect

Applied Acoustics

journal homepage: www.elsevier.com/locate/apacoust

Experiments on authenticity and plausibility of binaural reproduction via headphones employing different recording methods ☆

Josefa Oberem [a,*], Bruno Masiero [b], Janina Fels [a]

[a] Institute of Technical Acoustics, Medical Acoustics Group, RWTH Aachen University, Kopernikusstraße 5, 52074 Aachen, Germany
[b] University of Campinas, Av. Albert Einstein, 400, 13083-852 Campinas, SP, Brazil

ARTICLE INFO

Article history:
Received 7 January 2016
Received in revised form 8 July 2016
Accepted 12 July 2016

Keywords:
Binaural hearing
Authenticity
Plausibility
Individual HRTFs
Microphone setup

ABSTRACT

Major criteria for a successful binaural reproduction are not only a suitable localization performance, but also the authenticity and plausibility of the presented scene. It is therefore interesting to examine whether the binaural reproduction can be perceptually distinguished from a real source. The aim of the presented investigation is to compare the quality of the binaural reproduction via headphones with two different microphone setups (miniature microphone in Open-Dome and ear plug) for individual head-related-transfer-function (HRTF) and headphone-transfer-function (HpTF) measurements. Listening tests with a total of 80 subjects were carried out focusing on plausibility and authenticity. In the examination of plausibility detection rates showed that subjects were not able to match the reproduced pink noise to its reproduction system (real source vs. binaural reproduction via headphones). The authenticity of the static binaural reproduction was highly dependent on the stimulus. Pink noise could often be distinguished due to coloration in higher frequencies and small differences in location. A difference between microphone setups could not be found in neither of the listening tests.

1. Introduction

The idea of binaural recordings and reproduction has been explored from different points of view in various facets for several decades with profound results. However, binaural synthesis and reproduction, especially in practical application, can still be improved as it does not always yield perfect results. Therefore, this investigation focuses on the perceived quality of binaural reproductions.

Experiments in terms of listening tests are common for psycho-acoustic validations of binaural reproductions. Comparisons between real sources and binaural reproduction via headphones have been drawn in psycho-acoustic experiments especially regarding localization by for example Møller et al. [1], Wightman and Kistler [2] as well as Bronkhorst [3]. Investigations differed in stimulus type, duration, directions of sources, room conditions, headphone equalization and answering methods, among others. Overall a good agreement between results could be verified. Localizing with binaural reproduction was nearly as exact as localization with real sources for investigations for all three investigations [1–3].

Besides the demand of a physically correct reproduction and good localization, it is also important that the subject does not sense or hear a difference between real sources and the binaural reproduction. The indiscernibility between a binaural reproduction and a real source is a very high demand that can only be analyzed and proved in a direct comparison of the real and the virtual reproduction method. After Blauert [4], the perceptual identity is subsequently called authenticity. If a subject is only exposed to the binaural reproduction the perceptual identity is not essential, but it is sufficient if the subject rates the scene as plausible. Plausibility should be understood as "a simulation in agreement with the listener's expectation towards a corresponding real event" as defined by Lindau and Weinzierl [5]. Hence, for a plausible binaural reproduction the perceptual quality of the reproduction needs to be close enough to natural listening.

An early investigation on authenticity was carried out by Hartmann and Wittenberg [6]. In a listening test of forced choice design with four subjects they examined whether subjects were able to distinguish between the real source and the "virtual" binaural reproduction depending on a change of phase and level effects. Individual HRTFs were measured with probe microphones remaining inside the ear during the whole experiment. Using a synthesized vowel "a" as the stimulus the subject was asked to match

☆ Parts of this study were presented at the conferences: AIA-DAGA, Merano, Italy, 2013 and ICA, Montreal, Canada, 2013.

* Corresponding author.
E-mail addresses: job@akustik.rwth-aachen.de (J. Oberem), masiero@unicamp.br (B. Masiero), jfe@akustik.rwth-aachen.de (J. Fels).

http://dx.doi.org/10.1016/j.apacoust.2016.07.009
0003-682X/

the reproduction method to loudspeaker or binaural reproduction. Acoustically open headphones (*Sennheiser HD 40s*) were used to explore the perceived externalization.

Zahorik et al. [7] conducted a listening test with a 2-alternative-forced-choice design (2-AFC) to compare virtual and real sound sources with four experienced listeners. Individual HRTFs were measured with probe microphones for the binaural reproduction via supra-aural headphones (*Aiwa HP-M16*). Gaussian noise bursts (bandpassed: 300 Hz and 12 kHz) were presented from 15 different positions and results were analyzed as a function of filter length. Zahorik et al. [7] concluded that the virtual free-field was indistinguishable from the real free-field if a sufficiently long filter length was applied.

This result, however, was questioned by Langendijk and Bronkhorst [8] who carried out a listening test with a number of six listeners using a revised design to verify the results of Zahorik et al. [7]. Besides a 2-AFC design they also presented band limited noise bursts (500 Hz and 16 kHz) with an "oddball"-design and in a forced choice design (real vs. virtual) like Hartmann and Wittenberg [6] to examine the "fidelity of the three-dimensional-sound reproduction using a virtual auditory display" [8]. Detection rates were slightly but significantly above chance for the "oddball"-design. For the binaural synthesis HRTFs were measured with a probe microphone positioned at the eardrum and stimuli were played back by a midrange dome tweeter (*Sony MDR E-575*) mounted on a trolley.

One of the latest experiment on plausibility was carried out by Moore et al. [9] who tested the perceptual indistinguishability of a binaural reproduction using cross-talk-cancellation with eight subjects. The binaural synthesis was also based on individual data measured with probe-microphones in ear with one source position located in the frontal direction. In an "oddball"-design noise click-trains and harmonic pulses were presented yielding to the result that error rates were slightly but significantly underneath chance. Moore et al. [9] reported how perceived differences were due to an insufficient signal to noise ratio in high frequencies.

In another investigation published by Schärer and Lindau [10] in 2009 it was also analyzed whether binaural simulations could be perceptually distinguished from real sources. However, the main focus of this investigation was on seven headphone equalization methods and two different acoustically transparent headphones (*STAX SR5 2050II* and *STAX Lambda Pro New*) which were directly compared in a listening test with real sources. Most of the 28 subjects rated the binaural reproduction based on non-individual HRTFs as "boosting in high frequencies as well as ringing artifacts" [10]. The spectral coloration of the binaural simulation was also described as a major shortcoming by Lindau et al. [11]. Similarity rates between 0% and 70% were detected for pink noise and an acoustic guitar depending on the headphone equalization method. The authenticity of a binaural reverberant acoustical environment was tested in an ABC/HR-design.

Assuming that historical limitations of measuring techniques were the major reason for the use of probe microphones, it would be interesting to know whether a binaural reproduction measured with equipment that is state-of-the-art is comparably plausible. Difficulties as for example resonances in tubes and the notch filter effect are present in probe microphones and can be countered by new equipment. Different types of microphones used to measure HRTFs within the ear canal as well as the most adequate and applicable position in or around the ear have been investigated by several researchers [12–14]. Probe microphones were used by Wightman and Kistler [2] as well as Bronkhorst [3] among others due to size and signal to noise ratios, whereas in recent time measurements are more commonly made using miniature microphones placed at the entrance of the blocked ear canal [12,15]. In 1995, Møller et al. [16] measured HRTFs with an open auditory canal, but reported better results when HRTFs were measured with a blocked ear canal. However, the application and positioning of miniature microphones with silicon Open-Domes (cf. Section 2.1) is very simple, precise and little time consuming when HRTFs are frequently measured. Therefore, it could be asked whether the recording method (open meatus vs. blocked meatus) plays a significant role for the perception of the spatial sound reproduction.

Another technical aspect which should be taken into consideration are the headphones used for the binaural reproduction. Langendijk and Bronkhorst [8] criticized the headphones used by Hartmann and Wittenberg [6] as well as Zahorik et al. [7] and suggested the use of smaller headphones. In these investigations HRTFs were measured with headphones placed over the subjects' ears, resulting in deviations in higher frequencies, a spectral region known to contain important spectral localization cues [17]. For localization experiments this would surely be a major constraint, however, for the analysis of authenticity of a virtual sound source in a direct comparison, correct localization is not essential and HRTF measured with headphones could be used. However, the reproduction quality of the ear buds used by Langendijk and Bronkhorst should be questioned regarding transfer function and bandpass limitations. Acoustically open circumaural headphones (Sennheiser HD 600) were used in this investigation to reproduce binaural stimuli.

The demand for a plausible binaural reproduction is important in investigations where the binaural reproduction is only used as a tool and the aim of the analysis is to focus on other effects than localization or the perceived quality. Otherwise, experimental results will be biased. This is especially true for experiments that assess psychological effects like auditory attention [18–20] using binaural reproduction to simplify complicated laboratory situations. Frequently, individual HRTF are measured in different laboratories or at another time than the listening tests are conducted. Therefore, microphones need to be taken out of the ears and headphones are inevitably repositioned. In the present investigation HRTFs and HpTFs were measured separately as if measurement and experiment would have been taking place at a different place and time, even though the listening test was performed subsequently and in the same room.

The aim of this investigation was to examine the authenticity and plausibility of a binaural anechoic reproduction via open headphones depending on two different recording methods. In the method called "open meatus" a miniature microphone was positioned at the entrance of the open ear canal. "Blocked meatus" described the other method where the miniature microphone was placed into a foam closing the ear canal. Two listening tests were carried out. In the first listening test three different types of stimuli were used for a direct comparison of real sources and the binaural synthesis (authenticity). In a 3-alternative-forced-choice design (3-AFC) subjects were asked to find the stimulus which was different from the other two and therefore it was asked whether the subjects were able to distinguish between reproduction methods. In a second test pulsed pink noise was presented either by loudspeakers or as a binaural synthesis by headphones. Subjects were asked to define the reproduction method. In this indirect comparison the plausibility of the binaural reproduction was analyzed.

2. Methods and equipment

2.1. Microphones

To measure individual HRTFs and HpTFs, miniature microphones (*Sennheiser KE-3*, for the microphone's frequency response,

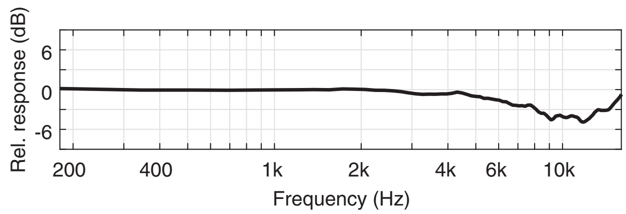

Fig. 1. Frequency response of a *Sennheiser KE-3* miniature microphone.

see Fig. 1) with a diameter of 3 mm were placed at the entrance of the participant's ear (cf. Fig. 2). Hammershøi and Møller [12] showed that the entrance of the ear canal was a suitable point for binaural recordings, since the further sound propagation towards the eardrum was independent of the direction of incidence.

In this investigation two recording methods were compared. For the first method (later called open meatus) the miniature microphone was fixed by a little silicon carrier called Open-Dome (cf. Fig. 3, to the left). Even though the silicon carrier did not close the ear canal, it needs to be mentioned that the microphone itself and the perforated carrier interfered with the entrance of the ear canal and therefore it was not completely open as under normal conditions, but could be described as partly open. Open-Domes come in different diameters and could therefore be conveniently fitted to the subjects meatus.

For the second method (later called blocked meatus) a commercial ear plug made of a damping foam was used to fix the microphone (cf. Fig. 3, to the right). The ear plugs were shortened in length for a comfortable fit and to ensure that the microphone was positioned in a way to be flush with the entrance of the ear canal (cf. [21]). With an ear plug the entrance of the ear canal was blocked.

Due to the anatomical variety of the ear canal entrances, it was difficult to specify the preciseness of the positioning of the microphone. However, it could be assured that the cavum conchae was never disturbed by the ear plug or the silicon Open-Dome and the deviation of the microphone's position from the anatomically defined entrance of the ear canal was less than 2 mm.

2.2. Headphones and loudspeakers

For the reproduction of the binaural synthesis open circumaural headphones (*Sennheiser HD 600*) were used, since they were acoustically transparent to exterior sound fields. Advantages over for example closed headphones were also depicted in findings by Møller et al. [22], Kleber and Vorländer [23] as well as Völk [15], who reported that open headphones usually show a coupling similar to the coupling to free air. Additionally, Kleber and Vorländer [23] carried out investigations on the impedance of different headphones. Findings showed that the headphone impedance was least influenced by movements for open headphones. Despite this, HpTFs of open headphones changed enormously with the positioning. Therefore, an adequate headphone equalization has to be used. A closer look on robust equalization methods was also taken by Masiero and Fels [24].

For the reproduction of real sources custom made coaxial loudspeakers were used. The frequency response varied within ±10 dB between 200 Hz and 20 kHz (cf. Fig. 4 for the frequency responses of all 24 loudspeakers). A compensation of the frequency response was applied individually for every loudspeaker and because of the challenges in low frequencies stimuli were bandpassed within the range of 200 Hz and 20 kHz.

2.3. Room setup

The listening tests as well as the HRTF measurements took place in a fully anechoic chamber ($l \times w \times h = 9.2 \times 6.2 \times 5.0\ \mathrm{m}^3$) with a lower boundary frequency limit of 200 Hz.

The subject was asked to sit inside a frame of 24 loudspeakers (cf. Fig. 5), which were equally distributed over azimuth in three elevation levels (0°, ±30°), whereas the distance was kept constant at 1.7 m. The chair was provided with a backrest, armrests and an adjustable headrest.

To control and minimize the movements of the subject's head an electromagnetic tracker (Polhemus Patriot[1]) was used during HRTF measurements and the listening test. Limits for the allowed head movements were set to ±0.5 cm in translation and ±2° in rotation.

Since individual filters were measured independently from the listening experiment, the subject was asked to sit in the same spot as during the measurements. The mounted headrest helped the participant to get back into the original position as well as instructions guided by the electromagnetic tracker within the defined range of translation and rotation.

Lights were turned off during the listening test to direct the focus to the aural sense [4].

2.4. Subjects

A number of 80 unpaid students and doctoral candidates aged between 20 and 36 who indicated normal-hearing, participated voluntarily in the experiment with a between-subject-design. All listeners, 40 of each sex, could be considered as non-expert listeners, since they were not trained in listening tests and were not familiar with the technique of binaural reproduction.

2.5. Binaural measurements, synthesis and equalization method

For the binaural synthesis in this investigation all HRTFs were measured individually for every subject. Measurements ran automatically with the *ITA-Toolbox* [25] in Matlab. Interleaved exponential sweeps[2] [26] were first sent to the sound card, then converted by an D/A-converter of type *Behringer ADA8000 Ultragain Pro-8* and amplified, and finally played back by the 24 loudspeakers in the anechoic chamber. The miniature microphone signal was pre-amplified, then went through the above-mentioned A/D-converter and the sound card before being post-processed (including time windowing). The signal-to-noise-ratio is about 80 dB in all measurements. Exemplary, lateral HRTFs (90°, to the participants left), measured with open and closed ear canal, are displayed in Fig. 6 in the frequency domain.

To examine the authenticity in a direct comparison of real sources and binaural reproduction headphones should stay on head during the whole listening test (cf. Section 3.2.1). Therefore, subjects also had to wear headphones during the HRTF measurement. It needs to be considered that the quality of localization suffered from this arrangement, but the examination of authenticity of real sources and binaural reproduction was of greater importance for this investigation.

In a second step HpTFs were measured to calculate an adequate robust equalization. After Masiero and Fels [24], headphones were repositioned on the subjects head after each of in total eight HpTF measurements. To give the best comfort, the repositioning was done by the subject itself. Based on Masiero and Fels [24] the

[1] Polhemus Patriot: Information given by the manufacturer: Update rate: 60 Hz; latency: 18.5 ms, static accuracy position: 1.5 mm, static accuracy orientation 0.4°.

[2] frequency range: 70 Hz–20 kHz, bit rate: 24 bit, sampling rate: 44.1 kHz, total excitation length: 7.5 s, no averaging.

Fig. 2. Miniature microphone placed at the entrance of the ear-canal using an Open-Dome (left) and an ear plug (right).

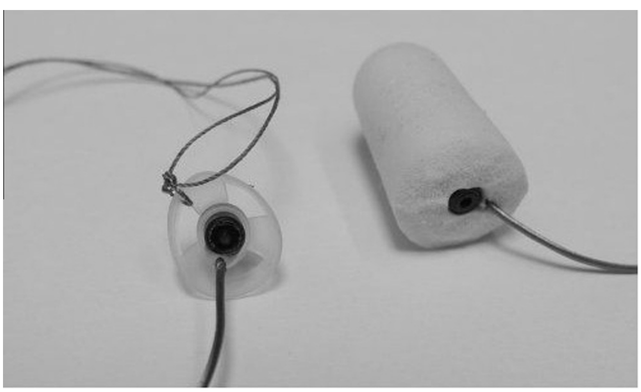

Fig. 3. Miniature microphone in Open-Dome (left) and in ear plug (right) to fix at the entrance of the ear canal.

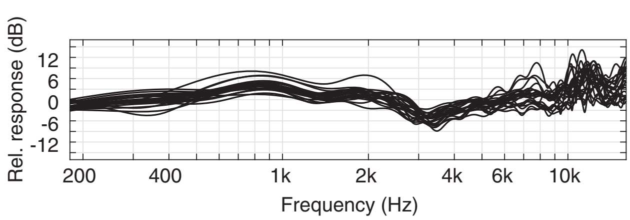

Fig. 4. Frequency response of the 24 loudspeakers. Since loudspeakers were custom made, transfer functions varied within ±10 dB.

equalization was calculated using the mean of the HpTF measurements. Since phase information was lost at this process, minimum phase was used. Furthermore, notches in the high frequency range were smoothed as particularized by Masiero and Fels [24]. Fig. 7 shows two single measurements of HpTFs with blocked and open meatus in frequency domain with an signal-to-noise-ratio of about 60 dB.

2.6. Additional measurements – real and virtual HRTF

To compare physically the frequency spectrum of the arriving sound produced by either real sources or headphones, reproducing a binaural synthesis, "real" and "virtual" HRTF measurements were performed. The measurements of "real" HRTFs were conform with the usual approach of HRTF measurements. To measure "virtual" HRTFs the binaurally synthesized stimulus was played by headphones and recorded with the microphone positioned at the entrance of the ear canal. To obtain a transfer function the recordings were divided by the original excitation signal. "Real" HRTFs and "virtual" HRTFs were compared in Fig. 8 for measurements with a blocked auditory canal and an open auditory canal. For perfect binaural reproductions the recorded signals were supposed to be identical.

The presented HRTFs were all measured from the same direction with the right ear. The source was positioned on the right with an elevation of $+30^\circ$. Overall a good agreement was given. Deviations in lower frequencies were due to windowing in the synthesis of binaural stimuli and did not exceed ±3 dB. Due to repositioning of headphones between measurements amplitudes for frequencies greater than 10 kHz could differ up to values about 10 dB, especially for an open auditory canal. In conclusion, it could be stated that both HRTFs, real and virtual, show a great similarity over the whole frequency range.

3. Experimental design

Two listening tests to examine the authenticity and the plausibility were carried out. Both experiments were based on an alternative forced choice design. In Experiment A a direct comparison was used to observe authenticity and in Experiment B the reproduction methods were compared indirectly to test the plausibility of the binaural reproduction.

A between-subject-design was chosen and as a consequence a number of 40 subjects form the group of subjects with HRTFs measured in blocked ear canal. Accordingly, the other 40 subjects belonged to the group of open meatus.

3.1. Stimuli

3.1.1. Experiment A: Authenticity

Three different stimuli were used in Experiment A. All three stimuli were bandpassed within the range of 200 Hz and 20 kHz. As mentioned before the range of the loudspeaker was limited to frequencies above 200 Hz. Inspired by Schärer and Lindau [10] pink noise and music were used. However, the 1.8 s lasting extract of a

Fig. 5. Anechoic room with loudspeaker setup and subject.

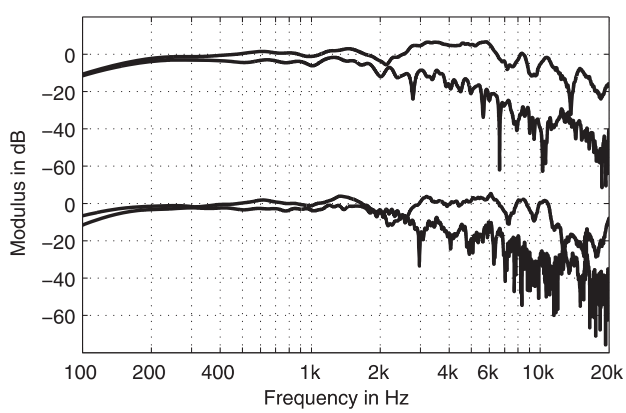

Fig. 6. HRTFs measured from the left (90°) with a blocked meatus (upper) and an open meatus (lower). For each pair the upper line depicts the left ear response and the lower line accordingly the right ear response.

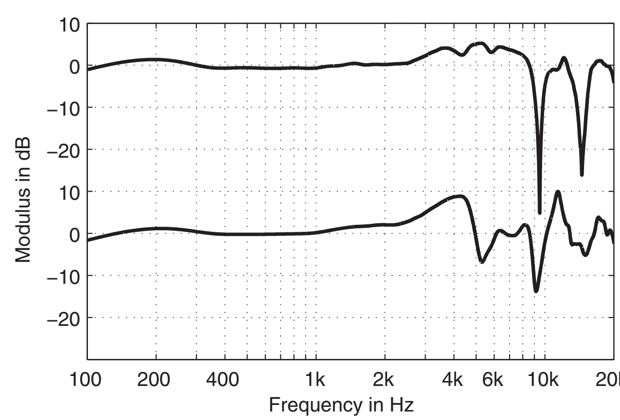

Fig. 7. HpTFs (of right ear) measured with a blocked meatus (upper) and an open meatus (lower).

music piece included also voices. The pink noise was pulsed since Rakerd and Hartmann [27] among others emphasized the importance of the onset of the stimulus. Its total length was 0.8 s, compound of to two pulses of 350 ms each (fade in/out: Hann window, 50 ms) and a break of 100 ms in between. The third stimulus used was an anechoic recording of the German word "Wunschdenken" spoken by a female containing three syllables, a "fizzing" sound and a sharp consonant. The duration was also 0.8 s.

All stimuli were convolved with the individually measured HRTFs and the headphone equalization for a binaural reproduction.

3.1.2. Experiment B: Plausibility

In Experiment B only one stimulus was used. The pulsed pink noise as described in Experiment A was presented in all trials. Again stimuli were convolved with HRTFs and the headphone equalization to reproduce binaural signals.

3.2. Experimental procedure

3.2.1. Experiment A: Authenticity

In a mixed design the two groups of open and blocked meatus were tested and compared regarding the reproduction method

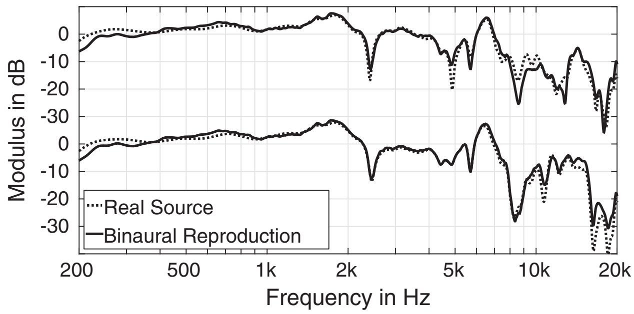

Fig. 8. Real and virtual HRTF measured with a blocked meatus (upper) and an open meatus (lower).

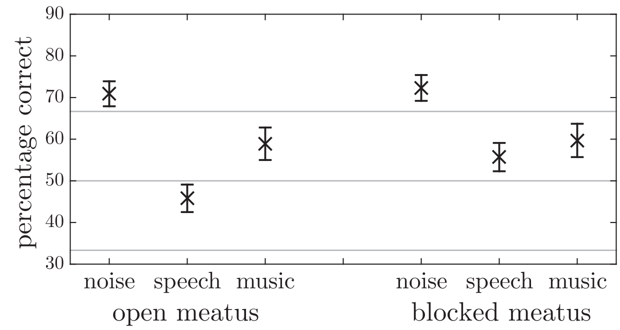

Fig. 9. Results for Experiment A – Authenticity of the binaural synthesis based on measurements with open and blocked ear canals against real sources in a 3-AFC design with three different stimuli. Error bars indicate standard errors.

(real sources vs. binaural reproduction) as well as the stimulus (noise vs. speech vs. music). Every subject completed one block including 20 trials of every stimulus. The authenticity was tested in a 3-AFC design for a direct comparison, where real sources and the binaural synthesis were presented immediately after another. Therefore, in one trial one stimulus (e.g. pink noise) was played three times in a row. Either one was played by loudspeakers (a), whereas the other two were binaurally reproduced by headphones (b), or the other way around (possible orders: aab, aba, baa, bba, bab, abb). The order of reproduction methods was randomly chosen and equally distributed over all subjects and over all directions. Moreover, playing levels were roved in 1 dB steps between 60 dB and 70 dB. Participants were asked to wear the headphones during the whole listening test, which made HRTF measurements including headphones necessary (cf. Section 2.5). Additionally, subjects were not told that the reproducing medium changes within one trial. Written instructions given to subjects at the beginning of each experiment only tell them to chose the sound out of three that differed without any specification. Moreover, subjects had the possibility to repeat a trial for a maximum of three times in case they had problems finding a difference. Written instructions, Play-Again-Button and buttons for the answer of a trial were given on a tablet computer. Therefore, subjects were able to carry out the experiment without any interference of the investigator.

3.2.2. Experiment B: Plausibility

After the subjects had completed Experiment A, they were asked to participate in Experiment B in the same session with a break of 5 min. Once again, a mixed design was used where the two previously described groups are compared regarding the reproduction method. This indirect comparison to analyze plausibility was based on a forced choice design with 10 trials for every subject, also randomized but equally distributed over all participants in direction of incidence and level. The pulsed noise was either played by the loudspeaker or as a binaural synthesis by headphones. Hence, different than in Experiment A, subjects listened to just one sound and had to answer whether the reproducing medium was a real source or the headphones. Like in Experiment A, participants worked with a tablet computer to enter their choices. Different than in Experiment A subjects did not have the chance to repeat a trial by pressing a Play-Again-Button.

4. Results

4.1. Experiment A: Authenticity

The results of Experiment A are shown in Fig. 9. The percentage of correctly answered trials were used to calculate means and standard errors split into different recording methods and stimuli. The data was submitted to a 2-way-ANOVA with the variables of recording method (R) and stimulus (S) depending on the authenticity of the binaural reproduction method. With respect to the chance of guessing and on account of the 3-AFC design a significantly smaller percentage of correct answers than 33.3% denotes that subjects were not able to hear any difference between the real source and the binaural reproduction in all trials.

Disregarding the kind of stimulus, subjects answered correctly and therefore heard a difference between reproduction methods in 58.5% of all trials when HRTFs and HpTFs were measured with an open meatus and in 62.6% of all trials when measured with a blocked meatus. The ANOVA yielded no significant main effect of recording method (R) regarding the reproduction method $F(1,78) = 1.92$, $p > 0.05$. Therefore, no significant difference between binaural synthesis based on HRTFs and HpTFs measured with open or blocked ear canals could be found.

The main effect of stimulus (S) regarding the reproduction method was significant $F(1,78) = 19.19$, $p < 0.001$. A post hoc t-test (LSD) was performed, with the outcome of significant differences between all stimuli. The interaction of recording method and stimulus ($R \times S$) regarding the reproduction method was not significant $F(1,78) = 1.12$, $p > 0.05$. For both recording methods subjects performed worst when the played stimulus was speech (open: 45.8%, blocked: 55.7%). For music (open: 58.9%, blocked: 59.7%) subjects also answered incorrectly within a high percentage of all trials, but the binaural reproduction seemed to be less authentic than the presentation of speech. Subjects had less difficulties to distinguish pink noise independent from the recording method (open: 70.9%, blocked: 72.3%). In subsequent surveys participants also stated how pink noise was easier to distinguish due to coloration in higher frequencies (40% of all subjects) as well as slight changes in location (76% of all subjects). The changes in location were also mentioned for the other stimuli presented (56% of all subjects). The opportunity to repeat a trial was frequently used. In 73% of all trials subjects listened for a second time and in 26% of all trials even for a third time.

4.2. Experiment B: Plausibility

The results of Experiment B are shown in Fig. 10. The percentage of correctly answered trials were used to calculate means and standard errors split into different recording methods. A Shapiro–Wilk test confirmed the normal distribution of the answers for each condition. The data was submitted to an ANOVA with the variable of recording method (R) depending on the authenticity of the binaural reproduction method. No significant difference between the reproduction methods could be found $F(1,78) = 2.33$, $p > 0.05$. All subjects of both groups had difficulties to distinguish between real sources and the binaural reproduc-

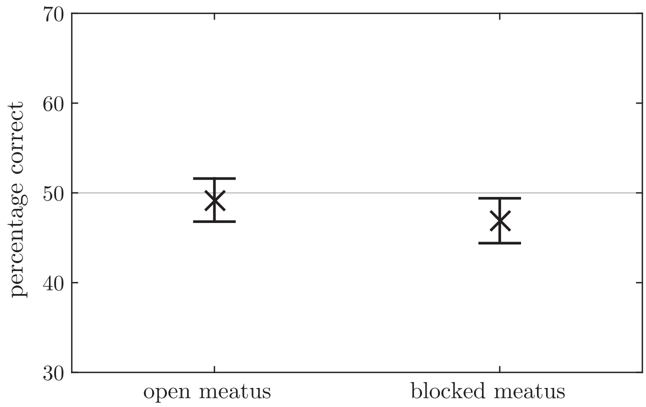

Fig. 10. Results for Experiment B – Plausibility of the binaural synthesis based on measurements with open and blocked ear canals against real sources in a forced choice design with pink noise. Error bars indicate standard errors.

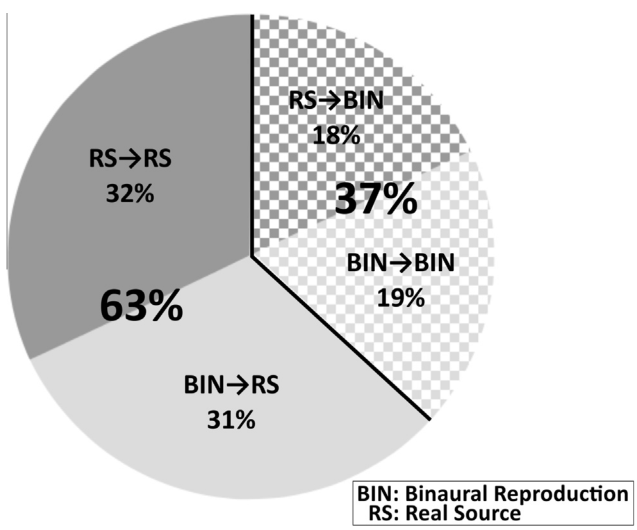

Fig. 11. Results Experiment B – Percentage of four combinations of playing reproduction methods and received reproduction methods. E.g. BIN → RS meaning binaural reproduction was delivered, however the subject selected that the Real Source played the noise.

tion in the indirect comparison, verified by the means of 49.3% and 46.9% especially with respect to the forced choice design. No significant difference from chance was found.

Fig. 11 shows that subjects chose rather real sources (63%) as the reproducing method than the binaural reproduction via headphones (37%), even though only half of the presented stimuli were delivered by loudspeakers. In a subsequent survey several subjects stated that they did not hear any difference between all trials and would have chosen the real source for 100%, but they felt uncertain since they also expected stimuli to be binaurally reproduced (85% of all subjects).

5. Discussion

The results of the conducted experiments showed that the individual binaural reproduction via acoustically open and individually equalized headphones was plausible for the applied recording methods. In Experiment B no significant difference from chance could be found. However, percentages of correct values were slightly smaller than 50% indicating that all subjects had difficulties to match real and virtual sound sources. On top of this, the majority of the listeners stated in the subsequent survey how they felt like all stimuli were presented from real sources.

Our findings were in agreement with the findings of previous studies. Hartmann and Wittenberg [6] reported how their subjects were not able to differentiate between real sources and the binaural reproduction when using a synthesized vowel as a stimulus. Furthermore, Zahorik et al. [7] described that listeners were not able to discriminate reproduction sources when noise bursts (bandwidth: 300 Hz–12 kHz) were presented in a 2-AFC design. Using a stimulus with a greater bandwidth (500 Hz–16 kHz), especially including higher frequencies like the stimulus used in this investigation (200 Hz–20 kHz) Langendijk and Bronkhorst [8] reported very similar detection rates for testing authenticity (forced choice design and 2-AFC).

In Experiment A, where the authenticity was analyzed, less convincing results were found. Differences between the real and the virtual source were in many cases clearly detectable for the listeners especially when pink noise was used as stimulus. A significantly smaller detection rate was found for music and speech. On account of the statements in the survey these difference were mainly due to differences in location. As described in the experimental setup (cf. Section 2.3) subjects were asked to move after having their HRTF and HpTF measured to simulate the usual procedure of a psycho-acoustical listening test whereby individual measurements and the listening test are conducted separately. Even though, they were guided back into the location of measurement small differences in location may appear due to the admitted variation around the original coordinates of the head. Similar to the results in Experiment B, the recording method supporting an open ear canal did not significantly differ from the method of a blocked meatus.

Plausibility was also analyzed by Langendijk and Bronkhorst [8], Moore et al. [9] as well as Schärer and Lindau [10]. Using non-individual HRTFs Schärer and Lindau obtained results differing between poor and satisfactory depending on the headphone equalization method. Subjects often perceived a boosting of high frequencies. Perfect authenticity between real sources and a non-individual binaural reproduction despite headphone equalization filters is almost inaccessible. Findings collected by Moore et al. were based on a binaural reproduction played back by loudspeakers using a CTC-filter. The detection rates were approximately between 45% and 60% depending on the type of stimulus. These results were remarkably good, since reproduction with CTC-filters often showed less encouraging results in localization or in psycho-acoustic issues than binaural reproduction via headphones [28,29,18]. The performance of the subjects in the present investigation was somehow better meaning, however, that the indistinguishability of real sources and virtual sources was worse than in Moore et al.'s experiment. Langendijk and Bronkhorst also found better results regarding plausibility in terms of detection rates with results only slightly significant from chance. Supposedly, differences between these investigation and the present experiment were due to the repositioning of the headphone as well of the subject inside the experimental room and the detachment of the microphones. Furthermore, subjects frequently used of the Play-Again-Button and therefore had the repeated chance to focus on small differences. It is unlikely that differences occurred because of the different microphone setups.

The main aim of this investigation was to compare recording methods (open meatus vs. blocked meatus) in terms of authenticity and plausibility. As predicted by Møller et al. [16], worse results in binaural reproduction based on HRTF measurements with an open meatus could have been expected. However, no significant difference could be found in this investigation. Results in terms of error rates/detection rates were nearly identical in both experiments. Therefore, no statistical proof for rejecting the null hypothesis is given. Proving a null hypothesis is difficult if not impossible. Nevertheless, it should not be neglected that in the performed experiments with a total of 80 subjects the recording methods in terms of authenticity and plausibility obtain nearly identical results.

6. Conclusion

The results of this investigation demonstrated that individual binaural reproduction with state-of-the-art methods in HRTF and HpTF measurements were overall plausible and therefore can be used in psycho-acoustical experiments or experiments assessing psychological effects like auditory attention in which HRTF and HpTF measurements and the listening test are conducted separately. A binaural reproduction via headphones being authentic is much more challenging than plausibility and were highly depended on the used type of stimulus. The authenticity obtained with the designed binaural reproduction could overall be rated as satisfactory. However, differences occurred due to the repositioning of subject and headphones. Using an adequate headphone equalization and a binaural synthesis the condition of the ear canal and the recording technique did not yield to different findings. The comfortable and little time consuming measuring method using Open-Domes can be recommended for HRTF and HpTF measurements in terms of plausibility.

Acknowledgments

The authors are grateful for the provided financing by DFG (Deutsche Forschungsgemeinschaft, Germany, FE1168/1-1).

References

[1] Møller H, Sørensen MF, Jensen CB, Hammershøi D. Binaural technique: do we need individual recordings? J Audio Eng Soc 1996;44:451–69.
[2] Wightman FL, Kistler DJ. Headphone simulation of free-field listening. II: psychophysical validation. J Acoust Soc Am 1989;85(2):868–78.
[3] Bronkhorst AW. Localization of real and virtual sound sources. J Acoust Soc Am 1995;98(5):2542–53.
[4] Blauert J. Spatial hearing – the psychophysics of human sound localization, vol. 373. USA-Cambridge MA: MIT Press; 1997. p. 36–50.
[5] Lindau A, Weinzierl S. Assessing the plausibility of virtual acoustic environments. Acta Acoust United Acoust 2012;98:804–10.
[6] Hartmann WM, Wittenberg A. On the externalization of sound images. J Acoust Soc Am 1996;99:3678–88.
[7] Zahorik P, Wightman FL, Kistler DJ. The fidelity of virtual auditory displays. J Acoust Soc Am 1996;99(4):2596(A).
[8] Langendijk EHA, Bronkhorst AW. Fidelity of three-dimensional-sound reproduction using a virtual auditory display. J Acoust Soc Am 2000;107 (1):528–37.
[9] Moore AH, Tew AI, Nicol R. An initial validation of individualized crosstalk cancellation filters for binaural perceptual experiments. J Audio Eng Soc 2010;58:36–45.
[10] Schärer Z. Lindau A. Evaluation of equalization methods for binaural signals. Audio eng conv, vol. 126. Munich, Germany, 2009. p. 7721.
[11] Lindau A. Hohn T. Weinzierl S. Binaural resynthesis for comparative studies of acoustical environments. In: Proc audio eng conv, vol. 122. Vienna, Austria, 2007. p. 7032.
[12] Hammershøi D, Møller H. Sound transmission to and within the human ear canal. J Acoust Soc Am 1996;100:408–27.
[13] Mehrgardt S, Mellert V. Transformation characteristics of the external human ear. J Acoust Soc Am 1977;61(6):1567–76.
[14] Middlebrooks JC, Makous JC, Green DM. Directional sensitivity of sound-pressure levels in the human ear canal. J Acoust Soc Am 1989;86(1):89–108.
[15] Völk F. Inter- and intra-individual variability in blocked auditory canal transfer functions of three circum-aural headphones. Audio eng conv, vol. 131. New York, USA, 2011. p. 8465.
[16] Møller H, Hammershøi D, Jensen CB, Sørensen MF. Transfer characteristics of headphones measured on human ears. J Audio Eng Soc 1995;43:203–17.
[17] Middlebrooks JC, Green DM. Sound localization by human listeners. Annu Rev Psychol 1991;42(1):135–59.
[18] Oberem J, Lawo V, Koch I, Fels J. Intentional switching in auditory selective attention: exploring different binaural reproduction methods in an anechoic chamber. Acta Acoust United Acoust 2014;100:1139–48.
[19] Fels J, Masiero B, Oberem J, Lawo V, Koch I. Performance of binaural technology for auditory selective attention. J Acoust Soc Am Hong Kong, China 2012;131 (4):3317(A). 13–19 May.
[20] Koch I, Koch V, Fels J, Vorländer M. Switching in the cocktail party: exploring intentional control of auditory selective attention. J Exp Psychol [Hum Percept] 2011;37(4):1140–7.
[21] Oberem J. Masiero B. Fels J. Authenticity and naturalness of binaural reproduction via headphones regarding different equalization methods. In: AIA-DAGA 2013: Proceedings of the international conference on acoustics Merano, Italy, 2013.
[22] Møller H, Jensen CB, Hammershøi D, Sørensen MF. Design criteria for headphones. J Audio Eng Soc 1995;43(4):218–32.
[23] Kleber J, Vorländer M. Messung von Gehöreingangsimpedanzen des freien Ohres und des abgeschlossenen Ohres mit Otoplastiken, Im-Ohr-Hörgeräten oder Kopfhörern (Measurements of impedances of the entrance of the open and blocked ear canal with Otoplastics, in-ear-hearing-aids and headphones). Hamburg, Germany: DGA; 2001.
[24] Masiero B. Fels J. Perceptually robust headphone equalization for binaural reproduction. In: Proc audio eng conv, vol. 130. London, UK, 2011. p. 8388.
[25] Institute of Technical Acoustics, RWTH Aachen: ITA-Toolbox. <http://www.ita-toolbox.org>; 2013.
[26] Dietrich P, Masiero B, Vorländer M. On the optimization of the multiple exponential sweep method. J Audio Eng Soc 2013;61:113–24.
[27] Rakerd B, Hartmann WM. Localization of sound in rooms, III: onset and duration effects. J Acoust Soc Am 1986;80:1695–706.
[28] Akeroyd MA, Chambers J, Bullock D, Palmer AR, Summerfield AQ, Nelson PA, et al. The binaural performance of a cross-talk cancellation system with matched or mismatched setup and playback acoustics. J Acoust Soc Am 2007;121:1056–69.
[29] Majdak P, Masiero B, Fels J. Sound localization in individualized and non-individualized crosstalk cancellation systems. J Acoust Soc Am 2013;133:2055–68.