# Effects of headphone transfer function scattering on sound perception

Mathieu Paquier, Vincent Koehl, Brice Jantzem

# EFFECTS OF HEADPHONE TRANSFER FUNCTION SCATTERING ON SOUND PERCEPTION

*Mathieu Paquier, Vincent Koehl*

University of Brest (UEB)
European Center for Virtual Reality (LISyC EA 3883)
25, rue Claude Chappe
29280 Plouzané, France

*Brice Jantzem*

Centre Auditif Entendre
4, square Commandant L'Herminier
29200 Brest, France

**ABSTRACT**

This study aims at evaluating the audibility of spectral modifications induced by slight but realistic changes in the headphone position over a listener's ears. Recordings have been performed on a dummy head on which 2 different headphone models were placed 8 times each. Music excerpts and pink noise were played over the headphones and recorded with microphones located at the entrance of the blocked ear canal. These recordings were then presented to expert and naïve listeners over a single test headphone. The subjects had to assess the recordings in a 3I3AFC task to discriminate between the different headphone positions. With the exception of one music excerpt for naïve listeners only, subjects were able to discriminate between the headphone positions.

## 1. INTRODUCTION

Sound reproduction over headphones is used in numerous applications such as sound quality assessment, binaural rendering and domestic use. Sound engineers often use headphones to monitor their recordings and mixes. When choosing a headphone model for a specific use, attention is paid to its type and especially to the quality of its transducers. Nevertheless, the coupling between the headphone and the listener's ears is not taken into account.

The HeadPhone Transfer Function (HPTF) describes both the headphone response and the coupling to a listener's ear. For binaural restitution (based on recordings or synthesis), the HPTFs can be measured, averaged (for repeated measurements) and inverted to compensate for the headphone influence and recreate the exact signals at the listener's ears. According to Pralong and Carlile [1], the equalization needs to be specific to the listener: they have measured the HPTFs of 10 subjects equipped with the same headphone by using an in-ear recording system. They found significant inter-individual differences in the 4 to 10 kHz range and showed that the use of non-individualized equalization can lead to errors in localization tasks.

Nevertheless, the signals being equalized or not - as it is the case for stereo recordings listened to over headphones and even for numerous cases of binaural restitution - the scattering caused by differences in the headphone position over the listeners' ears is not taken into account. However, it has been shown that slight modifications in the headphone placement can lead to large spectral differences. Toole [2] reported that these differences are less than 5 dB below 2 kHz, but ranged from 8 to 15 dB above 4-5 kHz. These differences were observed on 3 successive replacements on 3 human heads and on 3 dummy heads measured for 4 different types of headphones. Wightman and Kistler [3] and Pralong and Carlile [1] measured respectively the HPTFs on 10 humans for 10 headphone placements and on 10 humans and 1 manikin for 6 headphone placements. They reported that standard deviations of the magnitudes could reach up to 5 dB from 200 Hz to 14 kHz. McAnally and Martin [4] measured HPTFs for 20 headphone placements on 6 human heads. Standard deviations were generally smaller than 2.5 dB for frequencies up to 10 kHz, and be as high as 9 dB above 10 kHz. Kulkarni and Colburn [5] also observed a standard deviation of 9 dB on HPTFs measured for 20 headphone placements on an acoustic manikin for frequencies ranging from 9 to 14 kHz.

The perception of the sound scene might also be altered by HPTF variability if the localization cues are modified. The variability of the HPTF group delays being less than the minimum discriminable interaural time difference [4], the perceptual consequences of HPTFs variability would rather be provided by spectral differences. Kulkarni and Colburn [5] and McAnally and Martin [4] showed that HPTF and HRTF can exhibit similar spectral features. Martin et al. [6] have assessed the ability of listeners to localize sound presented via a virtual audio display that incorporated listener-specific equalization based on a single HPTF measurement. They found that listeners could localize virtual sound with free-field equivalent accuracy for eight headphone placements. So the headphone placement seemed to have a minor influence on this localization task. The variability observed on HPTF magnitudes (characterized by high-Q peaks and dips in high frequencies) is highly reduced when passing them through a cochlear filter model. McAnally and Martin [4] observed that the variability of the magnitudes of filtered HRTFs is generally considerably higher than that of the magnitudes of filtered HPTFs. This suggests that the spectral information used by listeners to localize sound is unlikely to be masked by the variability of the HPTF magnitude. However, even though the variability of HPTFs across headphone placements does not have an adverse effect on localization task, it could be perceived another way. As an example, for purposes of pure-tone audiometric testing, differences up to 15 dB have been observed in hearing thresholds because of bad headphone positioning [7]. Besides, a modification of the timbre could also be perceived.

The aim of this study is to evaluate whether realistic changes in the headphone placement can lead to audible changes in the sound perception. In this test, it is not possible to carry out a blind test by placing/removing the headphone over the listener's ears for the comparison of two consecutive placements. So, in the same way as for loudspeaker comparisons [8], the different headphone positions have to be recorded beforehand and played back over a fixed headphone. Three different monophonic sequences (one pink noise and two musical excerpts) were played over 2

different headphones and recorded with a dummy head. The omnidirectional microphones were located at the entrance of the blocked ear canal. The recorded sequences were then played back to expert and naïve listeners on a unique headphone for the whole test. These sequences were also filtered to compensate for the HPTFs of the test headphone. The listener's task was to compare recordings differing only in the headphone placement.

## 2. EXPERIMENTAL SETUP

### 2.1. Program material

Three short excerpts were used in this study. The first excerpt was a pink noise (3.5 s) and two music excerpts were selected from commercially available stereo material. They were extracted from CDs as 16-bit, 44.1-kHz Wave format files. The second excerpt (Ben Harper, 5 s) included drums, an acoustic guitar, a male human voice and choir voices. The third excerpt (Leonard Bernstein, 4 s) included a symphonic orchestra. As all of the 3 excerpts were presented in monophonic restitution, the left channels only were kept for the musical excerpts, mixing two recorded channels being generally not recommended [9].

### 2.2. Recordings

The recordings were made by using a dummy head (Neumann KU 100) whose microphones (omnidirectional) are located at the entrance of the blocked ear canal. Several studies have highlighted the benefits of this type of recording. From this point and up to the eardrum, the complete spatial information is included in the signal, but the inter-individual variability is much lower at the entrance because mainly caused by differences in the shape of the ear canal [11]. This technique also enables the use of rather large diaphragm microphones having better signal-to-noise ratios than the probe tubes used to measure the sound pressure within the ear canal [12]. Nevertheless, Møller et al. [10] put the emphasis on the fact that a blocked ear canal measurement is only valid in case of headphone with Free-air Equivalent Coupling to the ear (FEC) which means that the acoustical loading applied by the headphone on the ear canal is negligible. If the headphone does not fulfill the FEC conditions, which is the case for the closed headphones, the measure should be carried out in the open ear canal. The change in acoustic impedance occurring to the eardrum-ear canal system because of the ear closure by the headphone is thus taken into account. If a deviation of up to 4 dB in the Pressure Division Ratio (PDR: ratio between the pressures measured at the entrance of the open/blocked ear canal, with and without the headphone [10]) is tolerated, then most of the "open" headphones have FEC properties. Since the inter- and intra-individual differences observed on HPTF measurements are largely above 4 dB, this approximation is not excessive. In this paper, 2 different headphone models were under study:

- A: Sennheiser HD497 (supra-aural)
- B: Sony MDR CD580 (circum-aural)

These headphones were open and considered as having FEC properties. Each of them was placed and then removed over the dummy head by two different experimenters. The experimenters did not attempt to place the headphones in critical positions but only tried to reposition them as they should be normally placed over a listener's head. The 3 excerpts were thus recorded in 8 different positions for each headphone model.

### 2.3. Normalization of the recording levels

Even though recordings made on different headphones were not to be compared in the listening test, the mean levels per headphone were matched. So the possible differences in audibility of the positioning variability among the two headphones cannot be charged to differences in listening levels. The recording levels were thus normalized so that the average over the 8 placements is alike for the two headphones. This is equivalent to a compensation of the two headphone sensibilities. However, the relative levels of the 8 recordings made using one single headphone were not modified. The differences in these levels are caused by differences in the headphone position and shall not be compensated as they are a clue to perceive the positioning variability.

### 2.4. Test headphone equalization

The headphone used to perform all the subjective assessments was the Sony MDR CD2000, exhibiting particularly stable and broadband HPTFs. This headphone has been equalized by inverse filtering [10], when averaging over 8 HPTF measurements (Figure 1). These transfer functions have been obtained by placing the headphone over the dummy head. A sweep signal has been played over the headphone and recorded using the dummy head microphones. A headphone transfer function is generally characterized by smooth fluctuations in the low frequencies and by individual high-Q peaks and dips at high frequencies. An individualized headphone equalization can be achieved by considering the HPTFs measured on each subject. Although this kind of equalization seems preferable, an average equalization is acceptable and used in most cases [10]. Moreover, the goal of this study is the subjective comparison of different recordings for which the restitution artifacts are identically equalized. An exact and individual equalization is thus not mandatory.

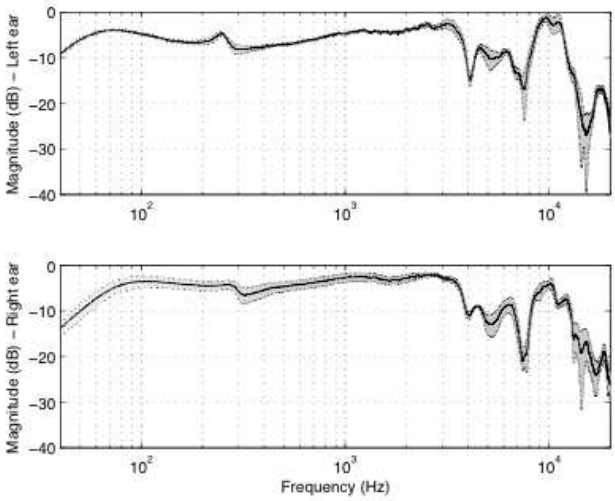


Figure 1: Transfer function (mean and standard deviation) of the Sony MDR CD2000 headphone for the left (up) and right ear (down).

### 2.5. Test procedure

The recordings made using the dummy head were presented during a listening test to determine whether differences in the headphone positions are noticeable. The test took place in an audiometric booth, the listener was sitting in front of a screen on which the answering buttons were displayed and could be activated by using a mouse. The stimuli were played back over the Sony MDR CD2000 headphone for the whole test. The subject was told to place it comfortably on his head and to not modify this position hence the test had started. The listener's task was to discriminate recordings differing only by the headphone placement using a 3I3AFC (3 Interval 3 Alternative Forced Choice) response paradigm. During a trial, 3 intervals were successively presented: one containing a certain recording position and two containing a different recording position. The recordings were randomly assigned to the intervals. The stimulus that was presented once (referred to as oddball stimulus) could then appear in first, second or third position. After having listened to the 3 stimuli (no repetition allowed), the subject had to indicate which one of the 3 intervals was the oddball one. The listener had to validate his choice to go on to the next trial. To compare this way all recordings made for one excerpt played over one headphone model (i.e. to compare the 8 positions), all possible combinations of 2 among 8 must be proposed (28 trials are needed to compare 8 stimuli). A "sub-session" was composed by all trials needed to compare all the recordings made using one single headphone model. Since 3 different excerpts were recorded for each headphone, a sub-session was made of 84 trials. During a sub-session, the listener had to assess the 28 trials associated with the first excerpt, then the 28 trials for the second one and finally the 28 trials for the third one. The excerpt order was randomized as well as the headphone model was selected at random for each sub-session.

A session lasted 45 min and was made of two sub-sessions, the listener having a 5-min break between them. The test was preceded by a 5-min pre-test to familiarize the listener with the answering interface and the stimuli. The listening level was identical for all listeners and was set to a realistic listening level for each excerpt, according to its content.

The listeners involved in this experiment were 10 "expert" and 10 "naïve" (without music or listening background) assessors according to the ISO 8586-2 standard [13].

## 3. RESULTS

### 3.1. Audibility of the positioning variability

A t-test indicated that the average detection rates (figure 2) were always significantly higher ($p<0.003$ in the least significant case) than 33.33% (equivalent to chance with a 3I3AFC response paradigm), except for the case "naïve listener, headphone A, excerpt 3 (Leonard Bernstein)", for which the mean detection rate (40%) was not significantly different ($p=0.1$) from chance level (33,33%).

### 3.2. Headphone effect

The analysis of variance showed that the headphone effect was significant ($F(1,108)=98.19$; $p<0.0001$). The detection task for the oddball stimulus appeared then to be significantly more difficult for the recordings made using the headphone A (Sennheiser HD497) than for the headphone B (Sony MDR CD580).

### 3.3. Listener's background effect

The analysis of variance showed that the listener's background effect was significant ($F(1,108)=47.39$; $p<0.0001$). The detection task for the oddball stimulus appeared then to be significantly more difficult for the naïve listeners than for the expert ones (Figure 2).

### 3.4. Excerpt effect

The analysis of variance showed that the excerpt effect was also significant ($F(1,108)=39.75$; $p<0.0001$). The Fisher LSD test indicated that the detection rate was significantly higher with the pink noise than with the two musical excerpts ($p<0.0001$). The two musical excerpts did not obtain significant different detection rates ($p=0.09$). So the detection task for the oddball stimulus appeared to be significantly easier for the recordings made using pink noise than music (Figure 2).

## 4. DISCUSSION

The most important result regarding this study is the confirmation that headphone positioning variability produces audible differences in most cases. The detection rate for the oddball sequence varies significantly across the headphones, excerpts and listener's background. The naïve listeners don't significantly detect the oddball stimulus with the Leonard Bernstein excerpt. However, for all the other cases, detection scores were always significantly higher to the value that could be obtained by chance. As it could be thought from past studies [4, 5], the action of placing and replacing a headphone over a listener's head causes significant modifications to the signal (objectively and perceptually). So the frequency smoothing applied by the inner ear [4] doesn't totally filter out the differences induced by successive replacements. In this study, the stimuli were obtained by recording the headphones using a dummy head, they were then placed by the experimenters themselves. It might be thought that the successive headphones replacements would have been more repeatable if carried out by the listeners themselves [10]. Nevertheless, the results significance, highlighting the obvious audibility of the differences caused by successive positions over a dummy head, could get to think that similar results would have been obtained for listener specific headphone positioning variability.

The audibility of the modifications induced by headphone positioning variability does not necessary signify that an equalization is mandatory to compensate for these differences. In binaural synthesis, it has been shown that the peaks characterizing the HPTFs were so high that the HPTF variability was negligible [4]. In the same way, it can be thought that the microphones, equalizations or treatments used by sound engineers modify the signal in a more significant manner than the differences in headphone placement. In addition, it would be quite hard to compensate for these differences: a real-time headphone equalization would require an in-situ (i.e. in the listener's ears) pressure measurement that would be by definition carried out in open ear canal conditions. This kind of measurement would be accomplished by using a probe tube and would therefore be of poor reliability

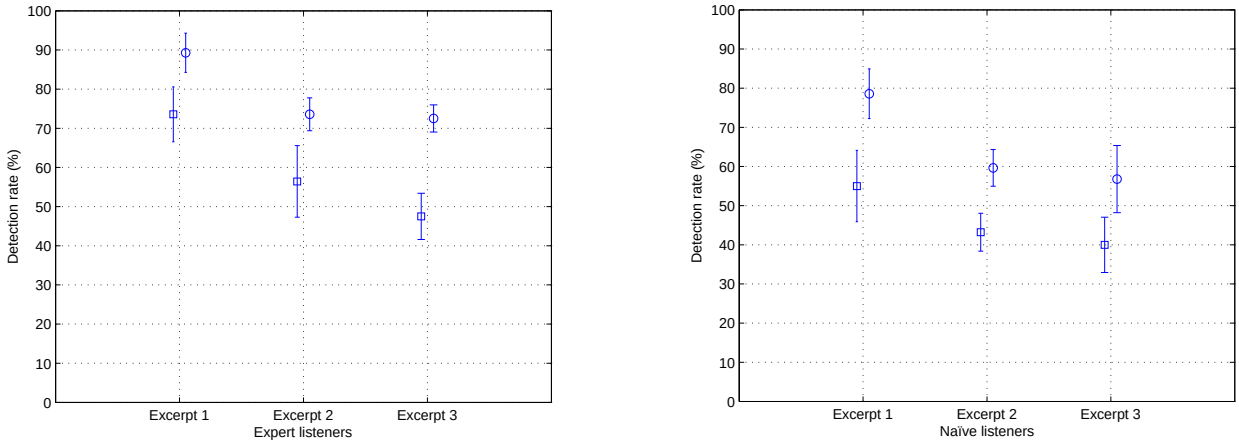


Figure 2: Mean detection rates for the headphones A (□) and B (○), the 3 excerpts, and the two listener groups within their 95% confidence interval.

because of the positioning accuracy, the frequency response of the probe microphones and the calibration difficulty.

In summary, the main conclusion that can be drawn from the results of this study would be a recommendation toward headphone users: the headphone positioning variability is highly perceptible and should be taken into account. The differences caused by variability in the headphone placement were more easily perceived using pink noise than musical excerpts. This statement may have several explanations: on one hand the pink noise is an almost steady-state signal which facilitates the memorization effort, which might be tough for musical excerpts, even very short. On the other hand, the spectral content of the pink noise ensure that the spectral modifications caused by a specific position, with possible high-Q peaks and dips, will be highlighted. This will not necessary be the case for music whose spectral content is much less regular and much more time-varying. Bücklein [14] observed that the detection of spectral peaks and dips was more accurate using white noise than music. He assumed that the audibility of the resonances increased with the spectral content.

## 5. CONCLUSION

Past papers indicated that spectral modifications are caused by headphone positioning variability. The present study showed that these spectral modifications led to audible differences in most cases. However expert and naïve listeners did not exhibit the same discrimination accuracy. The modifications caused by differences in the headphone position were always detected by expert listeners whereas they were not perceived by naïve listeners when the symphonic music excerpt was presented over one of the two different headphone models.

## 6. REFERENCES

[1] D. Pralong and S. Carlile, “The role of individualized headphone calibration for the generation of high fidelity virtual auditory space,” Journal of the Acoustical Society of America 100(1996), no. 6, 3785–3793.

[2] F.E. Toole, “The acoustics and psychoacoustics of headphones,” presented at the AES 2nd International Conference, Anaheim, CA, USA, 1984 May 11–14.

[3] F. L. Wightman and D.J. Kistler, “Headphone simulation of free-field listening. I: stimulus synthesis,” Journal of the Acoustical Society of America 85 (1989), no. 2, 858–867.

[4] K.I. McAnally and R.L. Martin , “Variability in the headphone-to-ear-canal transfer function,” Journal of the Audio Engineering Society 50(2002), no. 4, 263–266.

[5] A. Kulkarni and H.S. Colburn, “Variability in the characterization of the headphone transfer function,” Journal of the Acoustical Society of America 107 (2000), no. 2, 1071–1074.

[6] R. L. Martin, K. I. McAnally and M. A. Senova, “Free-field equivalent localization of virtual audio,” Journal of the Audio Engineering Society 49 (2001), no. 1/2, 14–22.

[7] D.S. Green, Pure-tone air-conduction threshold testing. In: J. Katz, Handbook of clinical audiology, fourth edition, Baltimore, MD: Williams & Wilkins, 1994; chapter 5, 97–108.

[8] F.E. Toole, “Binaural record/reproduction systems and their use in psychoacoustic investigations,” presented at the AES 91st convention, New York, NY, USA, 1991 October 4–8.

[9] AES20-1996 (Reaffirmed 2007), “AES recommended practice for professional audio – Subjective evaluation of loudspeakers,” Journal of the Audio Engineering Society 44 (1996), no. 5, 382–400.

[10] H. Møller, D. Hammershøi, C.B. Jensen and M.F. Sørensen, “Transfer characteristics of headphones measured on human ears,” Journal of the Audio Engineering Society 43 (1995), no. 4, 203–217.

[11] D. Hammershøi, H. Møller, “Sound transmission to and within the human ear canal,” Journal of the Acoustical Society of America 100 (1996), no. 1, 408–427.

[12] H. Møller, M.F. Sørensen, C.B. Jensen and D. Hammershøi, “Binaural technique: do we need individual recordings,” Journal of the Audio Engineering Society 44 (1996), no. 6, 451–469.

[13] ISO 8586-2, “Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 2: Experts,” International Organization for Standardization (2008).

[14] R. Bücklein, “The audibility of frequency response irregularities (1962),” reprinted in English in Journal of the Audio Engineering Society 29 (1981), no. 3, 126–131.