Discriminability of the placement of supra-aural and circumaural headphones

Mathieu Paquier, Vincent Koehl

Discriminability of the placement of supra-aural and circumaural headphones

Mathieu Paquier[*] and Vincent Koehl
University of Brest, Lab-STICC CNRS UMR 6285,
6 Avenue Le Gorgeu, 29238 Brest Cedex 3, France. mathieu.paquier@univ-brest.fr

Abstract— **Several studies have shown that differences in the placement of a headphone over a listener's ears could result in large differences in the measurements of the related transfer function (HPTF). Nevertheless, because of - at least - the non-uniform frequency resolution of human hearing system, large HPTF variations at some frequencies do not necessary imply audible consequences, which were not evaluated by past studies. The present study aims at evaluating the audibility of spectral modifications introduced by slight but realistic changes in the headphone placement over a listener's ears.**

Recordings were performed by placing/replacing a headphone on a dummy head. Various headphone models were realistically placed eight times each on the artificial head. Music excerpts and pink noise then were played back 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 discriminate between stimuli standing for different headphone placements using a 3I3AFC procedure.

With the exception of the naïve listeners about one given music excerpt only, subjects were always able to discriminate between the stimuli with respect to their corresponding headphone placement.

This indicates that consecutive realistic headphone placements may result in audible differences for the listener. Such a result could raise several issues about the use of headphones for psychoacoustic experiments, especially for multi-session tests.

Keywords—**Audibility, Headphone, placement.**

1. INTRODUCTION

Sound reproduction over headphones is used in numerous applications such as sound quality assessment, psychoacoustic experiments, domestic use, audio engineering, binaural rendering, etc [1,2]. When choosing a headphone model according to 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 apart from the fact that the user might prefer for example a circumaural open headphone over a supra-aural closed one.

The HeadPhone Transfer Function (HPTF) describes both the headphone response and the coupling to a listener's ear. For binaural rendering (based on recordings in ears or synthesis by convolving monophonic signals with HRTFs), the HPTFs can be measured, averaged (for repeated measurements) and inverted to compensate for the headphone influence so that the intended binaural signals are recreated at the listener's ear. According to Pralong and Carlile [3], the equalization needs to be specific to the listener. They found significant inter-individual differences in the 4 to 10 kHz range when measuring the HPTFs of 10 subjects equipped with the same headphone by using an in-ear recording system. In addition, they showed that the

use of non-individualized equalization can lead to errors in localization tasks. On the other hand, Lindau and Brinkmann [4] showed that non-individual binaural recordings were surprisingly perceived as most realistic when compensated using the HPTFs of the recording subject.

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 reproduction - the scattering caused by differences in the headphone placement 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, especially above 8000 Hz, where the quarter wavelength of sound is less than the length of the ear canal, and where standing wave patterns create large variations in sound pressure at different points within the canal. Moreover the large spectral difference of blocked-ear-canal HPTF at high frequency is caused by the difference in the coupling of headphone and pinna [5].

Toole [6] 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 for 3 successive headphone (4 different models) placements on real and dummy heads (3 each). Wightman and Kistler [7] measured the HPTFs on 10 humans for 10 headphone placements and Pralong and Carlile [3] did the same on 10 humans and 1 manikin for 6 headphone placements. Both studies reported that standard deviations of the magnitudes could reach up to 5 dB from 200 Hz to 14 kHz. McAnally and Martin [8] 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 were as high as 9 dB above 10 kHz. Kulkarni and Colburn [9] 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 as a result. The variability of the HPTF group delays being less than the minimum discriminable interaural time difference [9], the potentially audible modifications of HPTF would be exclusively spectral. Kulkarni and Colburn [9] as well as McAnally and Martin [8] showed that HPTF and HRTF can exhibit similar spectral features. Martin et al. [10] have assessed the ability of listeners to localize sound presented using a virtual audio display that enabled listener-specific equalization based on a single HPTF measurement. They showed that listeners were able to localize virtual sounds with an accuracy equivalent to free-field conditions for eight headphone placements. The headphone placement seemed thus to have a minor influence on this localization task. The variability observed in the HPTF magnitudes (characterized by high-Q peaks and dips in high frequencies) is highly reduced when applying a cochlear filter model on such frequency responses. McAnally and Martin [8] reported that the variability observed in the magnitudes of filtered HRTFs is generally considerably higher than the one observed in 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 still be perceived another way. As an example, since degraded or inadequate HRTF can enhance the "in head" sensation often evoked by headphone listeners [11,12], one can think that modifications of HPTF could have the same effect. As another example, headphone are often used for audiometric testing, and placement variability could possibly affect hearing threshold measurements, which was evidenced several past studies [13,14], but not by Gauz et al. [15]. Besides, at higher levels, a modification of the timbre could also be perceived because of resonances [16].

The aim of the present study is to evaluate whether realistic changes in the headphone placement (i.e. that could occur when anyone places a headphone on his own ears) can lead to noticeable changes in the sound perception. A blind test in which the listener would have to compare stimuli by placing/replacing the headphone over his own head is inherently impossible. So, in the same way as for loudspeaker comparisons [17,18], the different headphone placements have to be recorded beforehand and played back over a fixed

headphone. Three different monophonic sequences (one pink noise and two music excerpts) were played over 4 different headphone models 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 (experiment I with 4 recorded headphone models) and naïve (experiment II with 2 recorded headphone models) 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 because of the headphone placement at the recording step.

2. MATERIAL AND METHODS

2.1. Program material

Three short excerpts were used in this study. The first excerpt was a 3.5 s-long pink noise 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. These two music excerpts were chosen as symphonic music appeared to be more discriminant than popular music for resonance detection [19]. Moreover the masking effect is different with these two exerpts [20]. In order to allow the listener to concentrate on timbral modifications (and not on spatial ones) that could be introduced by differences in the headphone placement, monophonic excerpts were considered. Therefore, the left channels only were kept for the two music excerpts. The spectra of the two music excerpts are depicted in Fig. 1. The 3 excerpts were then diotically displayed to a dummy head using the various headphone models under test.

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, and there is less inter-individual variability when the recording point is the blocked entrance to the ear canal as compared to the open ear canal entrance and the eardrum because the blocked entrance does not include the effect of the ear canal [21]. 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 [22].

The headphone earpad is respectively surrounding or pressing the pinna for circumaural or supra-aural types. Usually, a circumaural headphone with a large cavity causes less compressive deformation of pinnae and therefore results in less HPTFs variation in different placements. In contrast, a circumaural headphone with a small cavity or supra-aural headphone inclines to cause large HPTFs variation in different placements [23]. In this paper, 4 different headphone models were under study:

· A: Sennheiser HD497 (supra-aural),
· B: Sony MDR CD580 (circumaural with cushions),
· C: Sennheiser HD600 (circumaural),
· D: Sony MDR CD2000 (circumaural with cushions).

These headphones are high-quality ones, especially models C and D which are commonly used by audio professionals and in psychoacoustic experiments. Headphones B and D are equipped with large earpads, surrounding the listener's pinna as for a circumaural model, but also with smooth cushions covering the pinna.
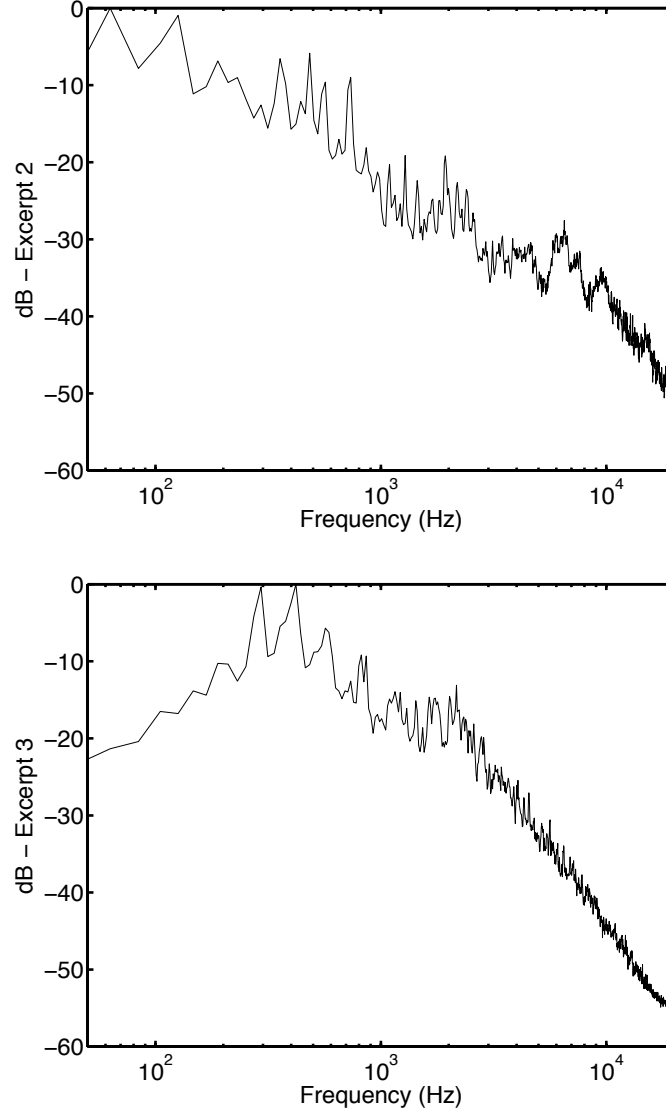
FIG. 1. Long-term spectra (left channel only) of excerpt 2 (Ben Harper) and excerpt 3 (Leonard Bernstein)

Each headphone 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 replace them as they should normally be over a listener's head. The 3 excerpts (plus a 20Hz-16kHz log-swept sine signal) were thus binaurally recorded for the 8 placements of each headphone model. The recordings were made using a laptop computer (MacBookPro) equipped with a MBox2 audio interface (f_s = 44100 Hz, 16 bits) and took place in an audiometric booth.

2.3. Normalization of the recording levels

Even though recordings stemming from different headphone types were not to be compared by the subject in the listening test, the mean levels per headphone models were matched. Differences in listening levels among the headphone models could possibly lead to discrepancies in the ability to discriminate between successive placements for a given headphone model. The recording levels were thus normalized so that the average over the 8 placements was alike for the four headphone models. This operation is equivalent to a compensation of the headphone sensitivities. 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 placement and shall not be compensated, as they could be a clue to perceive the placement variability.

2.4.Test headphone equalization

This part deals with the equalization of the unique headphone that was used for playback in the subjective test, denoted test headphone for the rest of the paper, and does not apply to the headphones when used for playback over the dummy head for recording purposes.

2.4.1. Equalization method

The headphone used to perform all the subjective assessments was the Sony MDR CD2000 exhibiting stable and broadband HPTFs (see Fig. 11). Moreover, this headphone obtained the highest sound quality ratings in an informal test conducted with sound engineers.

In order to enable the transparency of the test headphone, it was necessary to compensate for its transfer function An inverse filter needs to be computed according to:

$$H_{lin}(k) = H(k) \cdot H_{inv}(k) = 1 \qquad (1)$$

where k is the discrete frequency, $H(k)$ is the complex transfer function of the test headphone, $H_{lin}(k)$ is the linearized (i.e. compensated) transfer function of the test headphone and $H_{inv}(k)$ the complex inverse filter of the test headphone.

In the time domain, the convolution of h and h_{inv} (corresponding respectively to H and H_{inv}) should result in a system response h_{lin} equivalent to a dirac pulse:

$$h_{lin}(n) = h(n) * h_{inv}(n) = \delta(n) \qquad (2)$$

The direct inversion of the measured frequency response, i.e.

$$H_{inv}(k) = \frac{1}{H(k)} \qquad (3)$$

could however raise several problems. The direct inversion of mixed-phase systems yields an acausal, infinite and instable impulse response [24]. This can be avoided by splitting the transfer function into a minimum-phase and an allpass component, of which only the former is equalized [24]. This leads to an exact compensation of the amplitude but not of the phase response. However, in some cases the result of this method can be insufficient due to the remaining error energy in the allpass component [25,26]. This can especially cause the inverse filter to have large boosts at some frequencies, which could overload the system and could also produce a long filter. Several alternative approaches were developed: regularization [27], least square [28], homomorphic techniques [29], smoothing techniques (power spectra- or complex smoothing), with a fixed or a fractional-octave smoothing window [30,31].

For the present study, two inversion methods were experimented: the first one was a direct inversion of the minimum-phase component (because it is one of the simplest method to implement):

$$H_{inv}(k) = \frac{1}{H_{min}(k)} \qquad (4)$$

realized in the frequency domain (including zero padding).

The second method was a third-octave complex smoothing, which has proven to be perceptually superior in most cases according to Norcross et al. [32]. The principle is to reduce the severity of the peaks and dips by smoothing them prior to calculating the inverse.

In order to minimize the effect of the position of the test headphone in the inversion process, 8 HPTF measurements were carried by displaying a 5s-long log-swept sine (20 Hz-16 kHz) for each of the 8 headphone placements and by recording using the dummy head microphones (fe= 44100 Hz, 16 bits). Then these 8 HPTF were averaged (in the complex frequency domain). 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. Individualized headphone equalization could be achieved by measuring and considering the HPTFs related to each test subject. Although this kind of equalization seems preferable, an average equalization is acceptable and used in most cases [32]. 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.

A preliminary test was conducted in order to determine the perceptually best of these two inversion methods: recordings of the two music excerpts were convolved with the inverse response of the test headphone obtained with respect to the inversion method. Subjects had then to assess the quality of these compensated excerpts played back on the test headphone. This preliminary test indicated that the two methods resulted in equivalent quality ratings. The good ratings obtained using the direct inversion of the minimum-phase component can be explained by an already smooth appearance of the averaged response (over 8 positions) of the test headphone: the frequencies of the high-Q peaks and dips were slightly different for the successive positions, and the average over positions smoothed these differences. The use of a more sophisticated smoothing method, as the second one described above, was thus not necessary in our case.

As the two inversion methods appeared to have the same perceptual efficiency, the simple inversion of the minimum-phase component was chosen because it preserves better the test headphone HPTFs: Fig. 2 indicates that in high frequencies some recurrent evolutions of the curves for successive positions are preserved by the simple average, but removed by the smoothing. For low frequencies, smoothing curve is similar to the simple average.
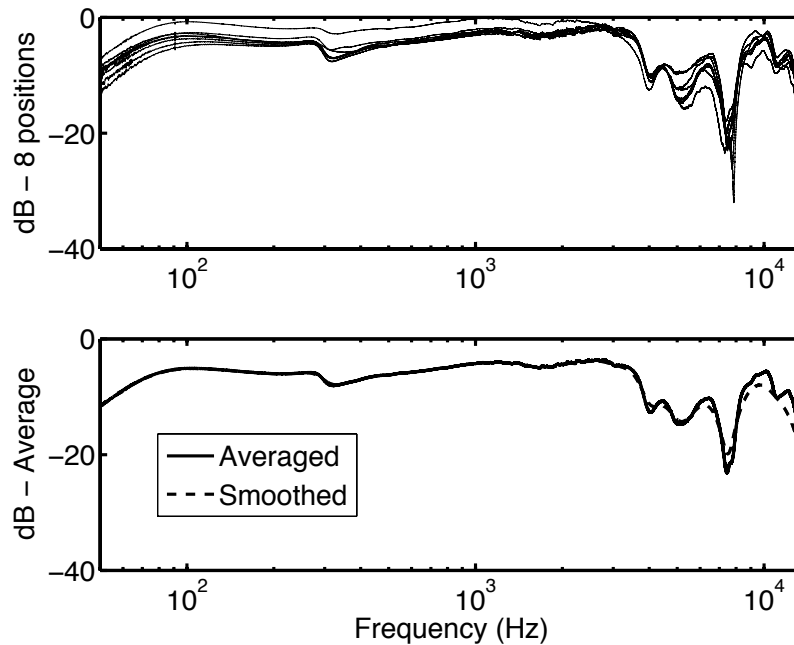

FIG. 2. HPTF modulus of the Test headphone (left ear), for each of the 8 measured positions (top), and averaged across the 8 positions, smoothed or not (bottom).

2.4.2. Test headphone coupling to the ear

When binaural recordings are carried out at the entrance of the blocked ear canal in order to be played back over a headphone, the coupling of the playback headphone has to be verified prior to inverting its response. According to Møller et al. [1,33], an additional equalization would be required for headphones that do not account for a Free-air Equivalent Coupling (FEC) to the ear.

The acoustical excitation from an headphone can be modeled by a Thevenin equivalent at the entrance to the ear canal [1]:

$$\frac{P_6}{P_5} = \frac{Z_{ear\,canal}}{Z_{ear\,canal} + Z_{headphone}} \qquad (7)$$

where:
• P_6 is the sound pressure at the entrance to the open ear canal,
• P_5 is the Thevenin pressure (does not normally exist physically, but if the ear canal is blocked as, for example, with an earplug, the Thevenin pressure is found at the outer side of the earplug),
• $Z_{earcanal}$ is the input acoustic impedance of the ear canal,
• $Z_{headphone}$ is the headphone radiation acoustic impedance seen from the ear canal.
according to the terminology of Møller et al. [1].

When the recording is made at the entrance to the blocked ear canal, it can be shown [1] that the correct eardrum signals are obtained during playback if the electrical gain G of the transmission is given by the expression:

$$G = \frac{1}{M.HPTF} \cdot \frac{\left[\frac{P_{open}}{P_{blocked}}\right]_{free\,air}}{\left[\frac{P_{open}}{P_{blocked}}\right]_{headphone}} \qquad (8)$$

where :
• M is the transfer function from pressure to voltage of the recording microphone,
• $HPTF$ is the electroacoustical transfer function of the headphone, measured at the entrance to the blocked ear canal,
• p_{open} is the sound pressure at the entrance of the open ear canal,
• $p_{blocked}$ is the sound pressure at the outer side of the blocked ear canal.

The last term of Eq. 8 can thus be defined as the Pressure Division Ratio (PDR). It can be expressed as follows when replacing $Z_{headphone}$ by $Z_{radiation}$ (free-air radiation acoustic impedance seen from the ear canal) in Eq. (7) for the free-air situation.

$$PDR = \frac{\left[\frac{P_{open}}{P_{blocked}}\right]_{free\,air}}{\left[\frac{P_{open}}{P_{blocked}}\right]_{headphone}} = \frac{Z_{earcanal} + Z_{headphone}}{Z_{earcanal} + Z_{radiation}} \qquad (9)$$

The two pressure divisions in Eq. (8) and Eq. (9) are measured respectively by displaying an excitation through a loudspeaker for free-air conditions or by the headphone itself. If the two pressure divisions are

equal, the *PDR* reduces to unity. A headphone complying to this statement is defined as FEC, and recordings carried out at the entrance of the blocked canal can be considered without any compensation due to *PDR*.

The FEC- compliance of the test headphone (Sony MDR CD2000) was verified by measuring P_{open} and $P_{blocked}$ using a 8 cm-long probe tube (ext. diameter 1.5 mm, int. diameter 0.5 mm) mounted on a DPA4060 miniature omnidirectional microphone with a coupler prototype. $P_{blocked\text{-}headphone}$ was measured on 9 ears by placing the free end of the probe tube at the entrance of blocked ear canals. Measurements were carried out on real heads for which the ear canal was blocked using individual moldings. The headphone was placed on the head by the subject himself. A log-swept sine was then played back over the headphone, and recorded with the probe microphone connected to a computer sound card. This operation was repeated for 8 headphone placements, the position of the probe tube being controlled by the experimenter between two consecutive measurements. A similar operation was carried out (with a repetition) in free air when reproducing the log-swept sine by using a loudspeaker (Genelec 8040) in order to measure $P_{blocked\text{-}free-air}$.

$P_{open\text{-}headphone}$ and $P_{open\text{-}free-air}$ were measured the same way once the individual moldings were removed.

The log-swept sine signal was used because this technique is acknowledged to provide good signal-to-noise ratio, and is nowadays preferred to other methods (linear sweep, MLS...) for measurement of the transfer function of weakly non-linear (or/and approximately time-invariant) systems [34].

FFT moduli were respectively averaged over the 8 headphone measurements and over the 2 loudspeaker measurements, and the *PDR* could then be calculated. Fig. 3 indicates the *PDR* means and standard deviations (computed on the log-magnitude of the *PDRs* for the 9 ears), as a function of frequency. The *PDRs* are approximately flat at frequencies below 2 kHz, and remain within the range of ±2 dB between 2 kHz and 5 kHz, even if deviation across the different ears is not negligible. According to Møller et al. [33], a headphone with such PDR values can be considered as FEC-compliant. Therefore, blocked ear canal measurements are suitable to invert the test headphone response.
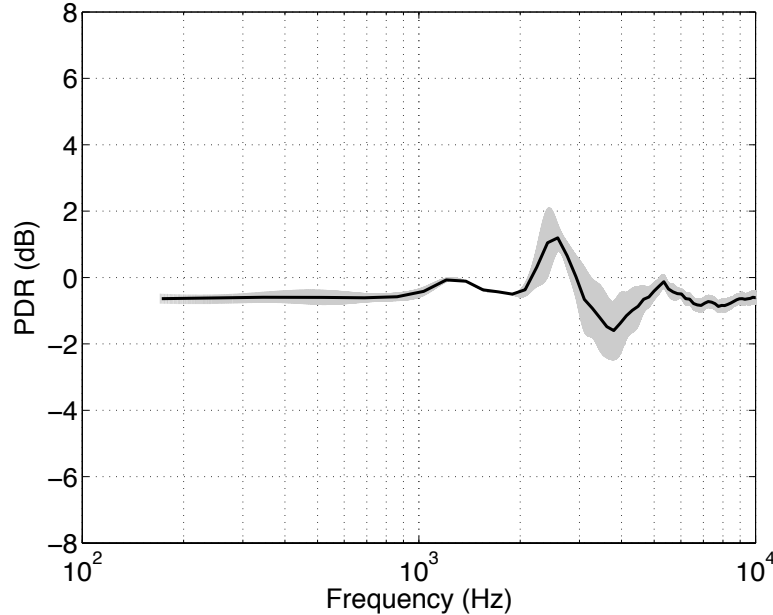

FIG. 3. Pressure Division Ratio : mean and standard deviation across the 9 ears.

2.5.Listening test procedure

Once compensated, the recordings made using the dummy head were presented during a listening test to determine whether differences due to headphone placement 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 dichotic stimuli were played back over the Sony MDR CD2000 headphone for the whole test. The subject was asked to place it comfortably on his head and to not modify this placement once the test had started.

The listener's task was to discriminate recordings differing only by the headphone placement using a 3 Interval 3 Alternative Forced Choice (3I3AFC) response paradigm. During a trial, 3 intervals were successively presented: one containing a given recording placement and two containing another recording placement. 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 placements), 28 trials (number of possible pair combinations to compare 8 stimuli) were proposed. A "sub-session" consisted of 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. The test was made of all sub-sessions needed to assess the headphones under test.

3. EXPERIMENT I

In order to assess the recordings stemming from the four headphone models, the test was run in two sessions lasting 45 min each. Each session was made of two sub-sessions, the listener having a 5-min break between them (he could remove his headphone during the break). The two sessions had to take place on two different days. The first of these two sessions 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, stimuli were played back at an average level of 85 dBSPL (according to a long term analysis of recordings realized when placing the test headphone over the dummy head, whose microphones were calibrated).

The listeners involved in the test were 10 sound engineering students from the "Master Image & Son" training program of the University of Brest. The subjects had no particular experience in listening tests. However, because of their experience, personal knowledge and interest, they were considered as competent for this listening test. According to ISO 8586–2 standard [35], they match the definition of "expert". The listeners were paid for their participation and showed normal hearing thresholds based on an audiogram [36] passed in the two months preceding the test.

3.1.Audibility of the placement variability

Using a 3I3AFC response paradigm, a subject answering at random has a one-in-three chance of pointing out the oddball stimulus. A detection rate for the oddball stimuli of 33.33 % would then be equivalent to chance. A t-test was used to determine whether the experimental detection rates were statistically higher than 33.33 %. The t-test indicated that, whatever the headphone or the excerpt, the average detection rates (Fig. 4) were always statistically different from 33.33 %, with $p = 0.001$ for the least significant cases.
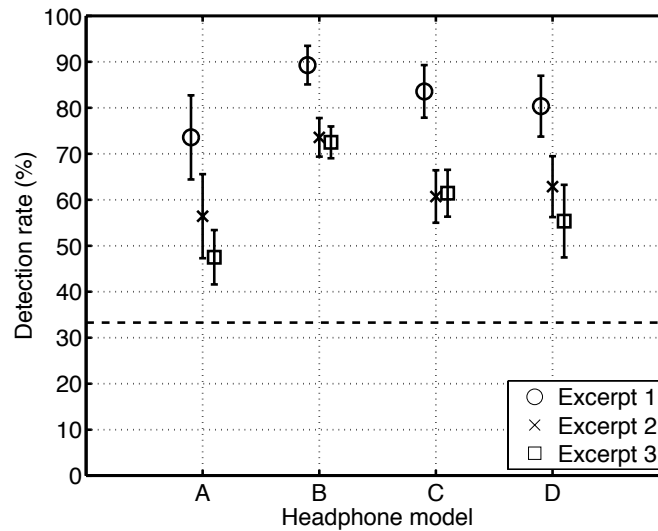
FIG. 4. Mean detection rates for the 4 headphones and the 3 excerpts, within their 95 % confidence intervals, when considering all expert listeners.

3.2.Headphone effect

As illustrated in Fig 5, a 2-way analysis of variance showed that the headphone effect was statistically significant (F(3,108) = 20.199; p < 0.0001). According to the Fisher LSD test, all headphones obtained statistically different detection rates from one another (p < 0.0001) except for the models C and D (p = 0.345). The detection task for the oddball stimulus appeared then to be statistically more difficult for the recordings made using the headphone A (Sennheiser HD497) than for the three other headphones. On the other hand the audibility of the placement variability was statistically higher for the recordings stemming from headphone B (Sony MDR CD580).
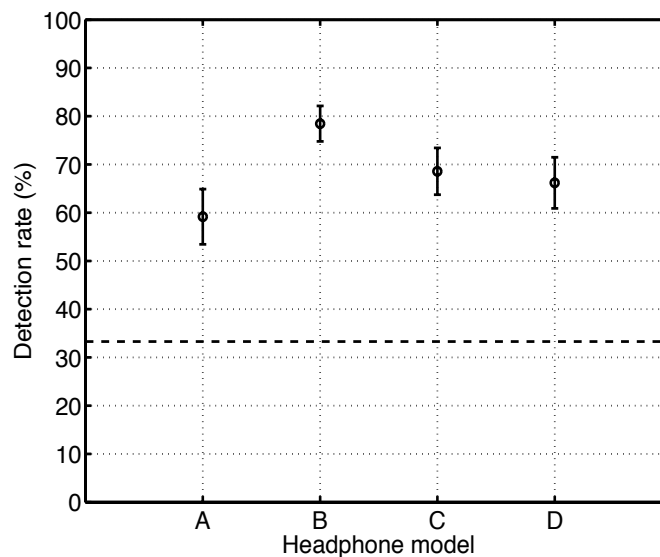

FIG. 5. Mean detection rates for the 4 headphones, within their 95 % confidence intervals, when considering all excerpts and all expert listeners.

3.3. Excerpt effect

As illustrated in Fig 6, the analysis of variance showed that the excerpt effect was also statistically significant ($F(2,108) = 60.603$; $p < 0.0001$). The Fisher LSD test indicated that the detection rate was statistically higher with the pink noise than with the two music excerpts ($p < 0.0001$). The two music excerpts did not obtain statistically different detection rates ($p = 0.056$). So the detection task for the oddball stimulus appeared to be statistically easier for the recordings made using pink noise than music.
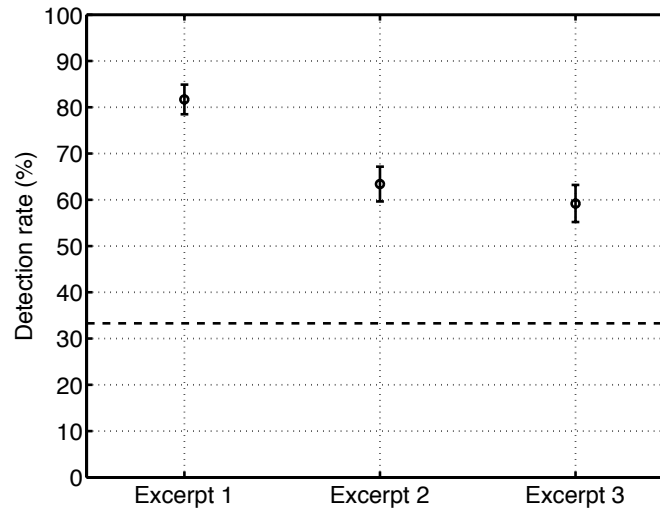

FIG. 6. Mean detection rates for the 3 excerpts, within their 95 % confidence intervals, when considering all headphone models and all expert listeners.

4. EXPERIMENT II

As a result to experiment I, the different headphone placements could be statistically distinguished when listening to the recordings carried out whatever the reproduction headphone model and the excerpt. However, it may be argued that such high detection rates could be obtained because of the expert status of the listeners involved in this experiment, who are used to critical listening. It can be thought that the results significance would not have been as high with naïve listeners. In order to investigate this hypothesis, the experiment was conducted with 10 naïve assessors (without music or listening background) according to ISO 8586–2 standard [35]. In order to reduce the test duration, only two headphone models were considered in this second experiment. The headphones under test were then reduced to models A (Sennheiser HD497, supra-aural) and B (Sony MDR580, circumaural with cushions) which respectively provided the worst detection rate and the best one. This second test thus consisted of two sub-sessions and was accomplished in a unique 45-min session proceded by a 5-min pre-test to familiarize the listener with the answering interface and the stimuli.

4.1. Audibility of the placement variability

t-tests indicated that for naïve listeners, the average detection rates (Fig. 7) were always statistically higher (p < 0.003 in the least significant case) than 33.33 % (equivalent to chance with a 3I3AFC response paradigm), except for the case headphone A, excerpt 3 (Leonard Bernstein), for which the mean detection rate (40 %) was not statistically different ($p = 0.1$) from chance level (33.33 %).
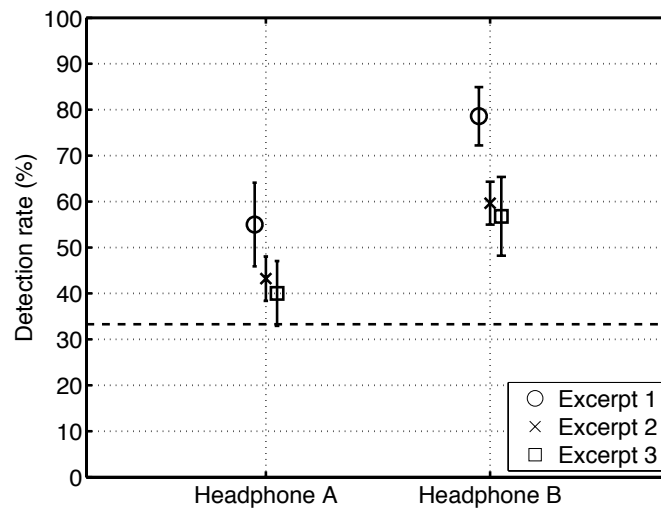
FIG. 7. Mean detection rates for the headphones A and B, and the 3 excerpts, within their 95 % confidence intervals, when considering all naïve listeners.

4.2. Headphone effect

As illustrated in Fig 8, a 2-way analysis of variance showed that the headphone effect was statistically significant ($F(1, 54) = 42.46$; $p < 0.0001$). The detection task for the oddball stimulus appeared then to be more difficult for the recordings made using the headphone A (Sennheiser HD497) than for the headphone B (Sony MDR CD580).
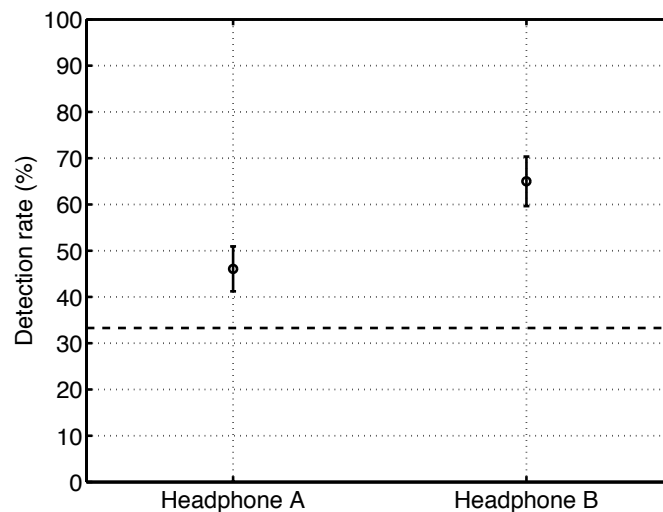

FIG. 8. Mean detection rates for headphones A and B, within their 95 % confidence intervals, when considering all excerpts and all naïve listeners.

4.3. Excerpt effect

The analysis of variance showed that the excerpt effect was also statistically significant ($F(2, 54) = 15.36$; $p < 0.0001$). The Fisher LSD test indicated that the detection rate was statistically higher with the pink noise than with the two music excerpts (p < 0.0001). The two music excerpts did not obtain statistically different detection rates ($p=0.09$). So the detection task for the oddball stimulus appeared to be statistically easier for the recordings made using pink noise than music (Fig. 9).
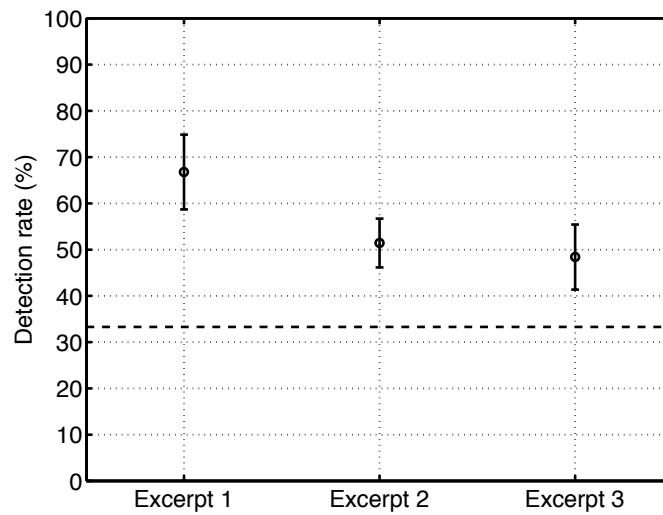
FIG. 9. Mean detection rates for the three excerpts, within their 95 % confidence intervals, when considering the headphones A and B and all naïve listeners.

4.4. Comparison with expert listeners

Headphone models A and B were tested by both the expert and naïve listeners. The experiments I and II could then be compared on the basis of the detection rates obtained by these two headphone models to look at the subject group effect. A 3-way mixed design analysis of variance (with "Headphone" and "Excerpt" as within-subject factor and "Subject group" as between-subject factor) was carried out to investigate the effects of the factors and their possible interactions on the detection rate.

The listener group proved to have a statistically significant effect ($F(1, 108) = 47.39$; $p < 0.0001$). As can be seen in Fig. 10, the detection rate appeared to be statistically higher for the expert subjects. As was already observed separately for the two listeners groups, the headphone model ($F(1, 108) = 98.19$; $p < 0.0001$) and the excerpt ($F(2, 108) = 39.75$; $p < 0.0001$) had statistically significant effects which respectively resulted in higher detection rates for headphone B and for pink noise.

In addition, no statistically significant interaction was highlighted between the subject group and the two other factors: the experts distinguished the headphone placements with a higher mean detection rate than the naïve listeners, whatever the headphone model and whatever the excerpt.
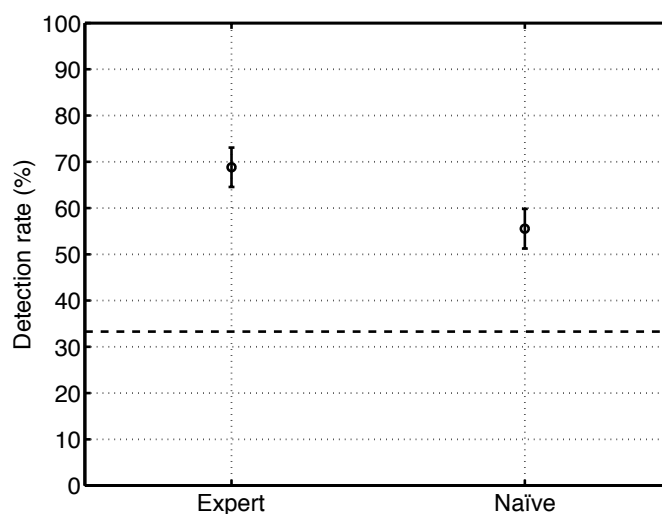

FIG. 10. Mean detection rates for expert and naïve listeners, within their 95% confidence intervals, when considering the headphones A and B and all excerpts.

5. DISCUSSION

5.1.Headphone placement is audible

The most important result regarding this study is that headphone placement variability produces audible differences in almost all cases. The detection rate for the oddball sequence varied significantly across the headphones, excerpts and listener's backgrounds but was statistically higher to the value that could be obtained by chance, except for one case only: the naïve listeners failed to detect the oddball stimulus using recordings carried out when reproducing the Leonard Bernstein excerpt with headphone A. As this headphone appeared to be the less discriminant in experiment I, it can be inferred that headphones C and D would have enabled discrimination for naïve listeners as well. Past studies showed that headphone placement appears to have no consequence on localization performances [8] nor - under certain conditions - on audiometry [15]. Nevertheless, as it could be suspected from other past studies [8,9], the action of placing and replacing a headphone over a listener's head have audible consequences. So the frequency smoothing applied by the inner ear [8] does not totally filter out the differences introduced by successive placements.

Fig. 11 indicates for the 4 headphones the means and standard deviations of non-normalized HPTFs across the 8 placements. These spectra were obtained from the 1024-pts magnitude spectrum of the impulse response obtained using a log-sweep sine signal recorded by using a miniature microphone located at the entrance of the ear canal (f_s = 44.1 kHz). These measurements were carried out on the dummy head at the same time as the recordings of the 3 excerpts that were used in the subjective experiment. In order to facilitate the reading, Fig. 12 indicates only the standard deviation per headphone model. These data are globally in agreement with previous studies [3,8,9,33] between which small differences can be observed in the measurements but might be due to the use of a different frequency step (in our study the 1024-pts FFT with f_s = 44.1 kHz provided a 43-Hz frequency step). For most of the headphones under test, the standard deviation remained below 2 dB for frequencies up to 5kHz, and reached up to 6 to 8 dB above 5 kHz.

According to more general studies about spectral peaks detection [18,37,38] and considering the frequencies, levels, and Qs of spectral peaks observed in the present HPTF measurements, the effects of headphone placement were presumably audible. This has been confirmed by the significance of the results to the present subjective experiments.

Finally, one should keep in mind that these results were obtained using monophonic excerpts (diotically presented to the dummy head for recording but resulting in dichotic signals because of the HPTFs). The possible extension of such conclusions to stereophonic or binaural recording must be taken with care.

5.2.Comparisons between headphones

The comparison of the measurements carried out on different headphones does not provide hints to explain the differences in audibility. As an example the placement variability was statistically better perceived by listeners on headphone B than on headphone A, whereas the standard deviation observed in the HPTF measurements appears to be higher for the headphone A than for the headphone B (in agreement with Riederer [23] who showed that the variability in HPTF measurements was higher for supra-aural headphones than for circumaural ones).

It is worth noting that the standard deviation reached up to 4 dB below 5 kHz on the right ear of headphone A, whereas it was only of 2 dB on the left ear, and on both ear for the other headphone models. This headphone was the cheapest of our selection, and it is possible that the right earpad was somehow misshapen, which was however not visually detectable. Nevertheless, despite such a high physical variability in "low" frequencies, the perceptual effect of placement was lower on headphone A.

For headphone B (the one which provided the easiest discrimination of placements), the standard deviation was particularly small above 5 kHz (lower than 4 dB), except for an isolated 7 dB peak at 8 kHz on the left ear. The ability to detect the effect of placement for this headphone could possibly be explained by this peak only. Finally, one should note that headphone A (Sennheiser HD497), which provided the worst detection rate in the perceptual test, and exhibited thus the highest robustness to placement

variability, was a supra-aural headphone. This finding is not in agreement with ANSI S3.21 standard [34] which states that the use of a circumaural headphone should increase the accuracy of auditory threshold determinations at high frequencies, nor with Riederer [23] and Atherley [14], who suggested the use of circumaural heaphones, which generally provide a better placement stability. Nevertheless, in these studies, the authors did not use the same headphones as those chosen in the present study, and the generalization of their finding has to be taken with care. In particular, the headphone A was clearly of a least quality than the three other ones. The least quality of this headphone is not obvious when looking at the HPTFs (Fig. 11), but was clearly audible by any listener. In addition, this headphone was significantly cheaper than the other ones. As the audibility of headphone placement would be exclusively caused by spectral cues [8], one can think that the headphone placement, characterized by modifications in high frequencies, would not be so critical for an headphone providing a bad reproduction of the spectrum, especially in high frequencies. On the other hand, the placement could be a more serious issue for higher quality headphones, enabling a more precise reproduction of the whole audible band.
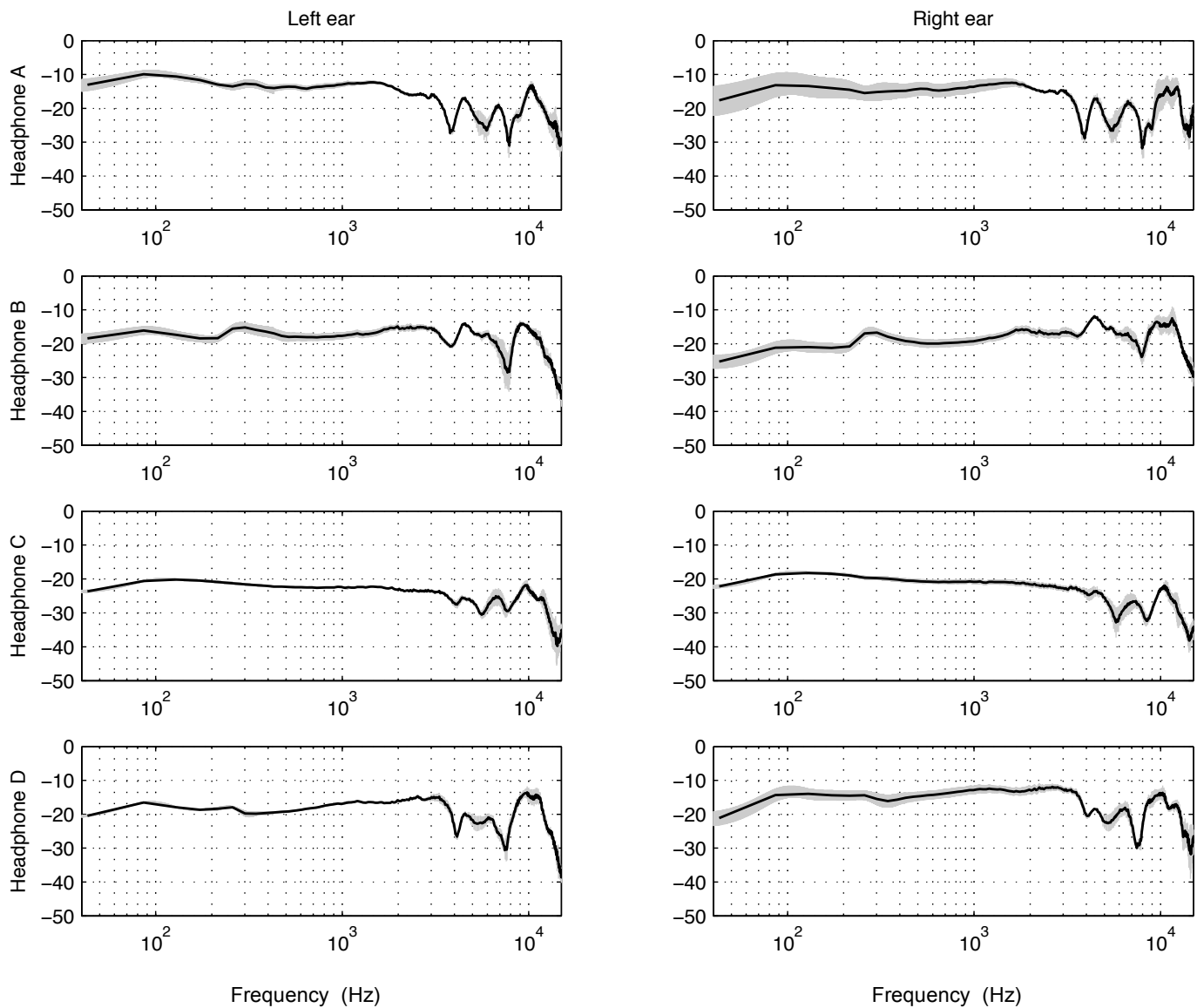

FIG. 11. Means and standard deviations of the HPTFs of 4 headphones (left and right ears, not normalized) computed on a decibel basis from 8 placements measured at blocked ear canal. Shaded areas mean ±1 standard deviation.
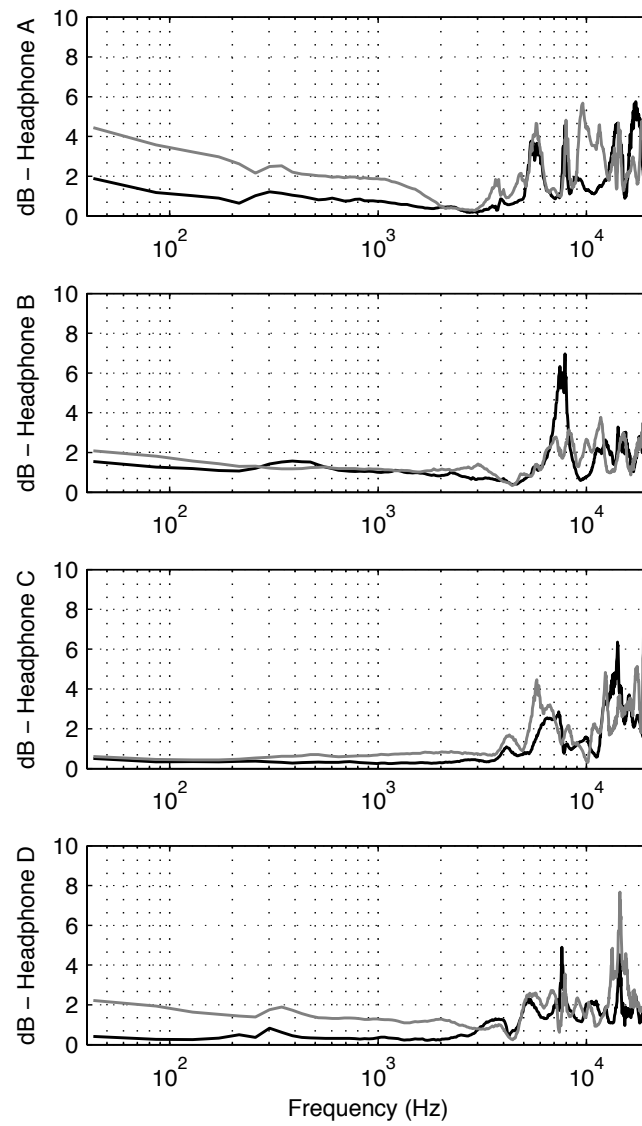
FIG. 12. Standard deviations of the HPTFs of 4 headphones (left ear in black, right ear in grey) computed on a decibel basis from 8 placements measured at blocked ear canal.

5.3.Excerpt influence

The differences caused by variability in the headphone placement were more easily perceived using pink noise than music excerpts. This statement may have several explanations: the pink noise is an almost steady-state signal, which facilitates a memorization effort that might be tough for music excerpts, even for very short ones. In addition, the spectral content of the pink noise ensures that the spectral modifications caused by a specific placement (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 [37] and Fryer [19] observed that the detection of resonances was more accurate using white noise than music, symphonic music being more discriminant than popular music. According to them, the broader the spectral content, the better the audibility of the resonances. Toole and Olive [16] have also observed that resonance detection at low frequencies was easier using pink noise than using music or speech signals and depended on the temporal behavior of the signals under test.

In this study, only the left channel of music recordings was kept. This could cause a bias toward the high frequencies for the third excerpt, as the left channel of symphonic recordings usually contains the first and second violin groups and substantially high-pitched woodwinds like flute and clarinet. As can be noted in Fig. 1, the long-term spectrum of this excerpt mainly contains frequencies above 300 Hz. Using a monophonic recording or an average of the two channels would have possibly given different results because of the frequency content and the possible masking effects. In addition, it is possible that stereophonic recordings would also reduce discriminability of HPTF differences due to positioning.

5.4.Consideration of headphone placement

The audibility of the modifications introduced by headphone placement variability does not necessary mean that an equalization is mandatory to compensate for these differences. Several applications do not require such compensation: for instance, in binaural synthesis, it has been shown that the peaks characterizing the HRTFs were so high that the HPTF variability was negligible for sound source location [8]. However, in psychoacoustic experiments where the listener may have to assess the frequency content, variability due to headphone placement might be an issue. As an example, inter-subject variability observed in a listening test (conducted with headphones) could be partially caused by inter-subject differences in HPTFs and might therefore be increased by odd headphone placements. In addition, for experiments conducted in several sessions and for which the headphone is placed several times on a given listener's head, placement variability could increase intra-subject variability.

Finally, it would be quite hard to compensate for HPTF 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 because of the positioning accuracy of the tube, the frequency response of the probe microphones [41], the difficulty of calibrating them (particularly above 7-8 kHz) and especially the bad signal to noise ratio of these devices [17].

A possible solution could be the use of earphones rather than classic headphones. Earphones were not considered in this study because the microphones in use in the present study were located at the entrance of the blocked ear canal of the dummy head. To record the response of earphones, probe tubes should be used, thus involving the same drawbacks as mentionned above. With earphones, the influence of the pinna is not taken into account, and so it may be thought that variability of narrow peaks and dips would disappear. However the resonance of ear canal could change with the acoustic loading resulting from the earphone insertion, providing another form of HPTF variability, possibly audible as well.

6. CONCLUSION

Previous studies indicated that spectral modifications were caused by headphone placement variability. The goal of the present experiment aimed at indicating whether these spectral modifications were audible. At the exception of one condition (a given headphone, and only for one given excerpt and for naïve listeners), spectral modifications led to audible differences. The perception of these differences can be more or less emphasized according to the headphone model, the listener background, and the excerpt.

These audible differences could be of crucial importance in some applications, especially for psychoacoustic experiments conducted over several sessions in which a headphone has to be placed several times over subjects' heads.

To sum up, the main conclusion that can be drawn from the results of this study would be a recommendation towards auditory researchers working with headphones: spectral variability resulting from headphone displacement can be easily discriminated and can possibly bias experiments.

REFERENCES

[1] H. Møller, "Fundamentals of binaural technology", *Appl. Acoust*, vol. 36, pp. 171–218, 1992.

[2] H.W. Gierlich, "The application of binaural technology," *Appl. Acoust.*, vol. 36, pp. 219-243, 1992.

[3] D. Pralong and S. Carlile, "The role of individualized headphone calibration for the generation of high fidelity virtual auditory space," *J. Acoust. Soc. Am.*, vol. 100, pp. 3785–3793, 1996.

[4] A. Lindau and F. Brinkmann, "Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings," *J. Audio Eng. Soc.*, vol. 60, No 1/2, pp. 54–62, 2012.

[5] B. Zhou and D. Green, "Reliability of pure-tone thresholds at high frequencies," *J. Acoust. Soc. Am.*, vol. 98, pp. 828–836, 1995.

[6] F. Toole, "The acoustics and psychoacoustics of headphones," in *Proc. 2nd Audio Engineering International Conference*, Anaheim, CA, USA, 1984.

[7] F.L. Wightman and D. Kistler, "Headphone simulation of free-field listening. I: Stimulus synthesis, " *J. Acoust. Soc. Am.,* vol. 85, pp. 858–867, 1989.

[8] K. McAnally and R. Martin, "Variability in the headphone-to-ear-canal transfer function," *J. Audio Eng. Soc.*, vol. 50, pp. 263–266, 2002.

[9] A. Kulkarni and H. Colburn, "Variability in the characterization of the headphone transfer function," *J. Acoust. Soc. Am.*, vol. 107, pp. 1071–1074, 2000.

[10] R.L. Martin, K.I. McAnally, and M.A. Senova, "Free-field equivalent localization of virtual audio," *J. Audio Eng. Soc.*, vol. 49, pp. 14–22, 2001.

[11] L. Wang, F. Yin and Z. Chen, "An "out of head" sound field enhancement system for headphone," in *Proc. Neural Networks and Signal Processing*, Zhenjiang, China, june 8-10, pp. 517-404, 2008.

[12] J. Breebaart and E. Schuijers, "Phantom materialization: A novel method to enhance stereo audio reproduction on headphones", *IEEE Trans. on audio, speech and language proc.*, vol. 16, No 8, November 2008.

[13] J. Steinberg and W. Munson, "Deviations in the loudness judgments of 100 people," *J. Acoust. Soc. Am.*, vol. 8, pp. 71–80, 1936.

[14] G. R. C Atherley, T. I. Hempstock, P. Lord and J. G. Walker, "Reliability of auditory threshold determinations using a circumaural-earphone assembly," J. Acoust. Soc. Am., vol. 42, pp. 199–203, 1967.

[15] M.T. Gauz, D.O. Robinson and G.M. Peters, "High-frequency Békésy audiometry: III. Reliability and validity revisited", *J. Aud. Res.*, vol. 2, pp. 167–80, 1981).

[16] F. Toole, and S. Olive, "The modification of timbre by resonances: Perception and measurement", *J. Audio Eng. Soc.*, vol. 36, pp. 122–142, 1988.

[17] F. Toole, "Binaural record/reproduction systems and their use in psychoacoustic investigations," in *Proc. 91st Audio Engineering Society Convention*, New York, NY, 1991.

[18] T. Hiekkanen, A. Mäkivirta, and M. Karjalainen, "Virtualized listening tests for loudspeakers," *J. Audio Eng. Soc.*, vol. 57, pp. 237–251, 2009.

[19] P. Fryer, "Intermodulation distortion listening tests," in *Proc. 50th Audio Engineering Society Convention*, London, UK, 1975.

[20] V. Koehl and M. Paquier, "A comparative study on different assessment procedures applied to loudspeaker sound quality", *Appl. Acoust*, vol. 74, pp. 1448–1457, 2013.

[21] D. Hammershøi, H. Møller, "Sound transmission to and within the human ear canal," *J. Acoust. Soc. Am.*, vol. 100, pp. 408–427, 1996.

[22] H. Møller, M. Sørensen, C. Jensen and D. Hammershøi, "Binaural technique: Do we need individual recordings," J. Audio Eng. Soc., vol. 44, pp. 451–469, 1996.

[23] K. Riederer, "Repeatability analysis of head-related transfer function measurements", in *Proc. 105th Audio Engineering Society Convention*, San Francisco, CA, 1998.

[24] J. N. Mourjopoulos, "Digital equalization of room acoustics, "*J. Audio Eng. Soc.*, vol. 42, pp. 884-900, 1994.

[25] M. Karjalainen, T. Paatero, J. N. Mourjopoulos, P. D. Hatziantoniou "About room response equalization and dereverberation", *IEEE Workshop on ASPAA*, New Paltz/NY, October, 2005.

[26] P. A. Nelson, F. Orduna-Bustamante, and H. Hamada, "Inverse Filter Design and Equalization Zones in Multichannel Sound Reproduction," *IEEE Trans. Speech and Audio Process.*, vol. 3, pp. 185–192, 1995.

[27] O. Kirkeby, P. A. Nelson, H. Hamada, and F. Orduna-Bustamante, "Fast Deconvolution of Multichannel Systems Using Regularization," *IEEE Trans. Speech and Audio Process.*, vol. 6, pp. 189–194, 1998.

[28] O. Kirkeby and P. A. Nelson, "Digital Filter Design for Inversion Problems in Sound Reproduction, *J. Audio Eng. Soc.,* vol. 47, pp. 583–595, 1999.

[29] J. N. Mourjopoulos, P. M. Clarkson, J. K. Hammond "A comparative study of least-squares and homomorphic techniques for the inversion of mixed phase signals" In Proc. of *IEEE Int. Conference on ASSP*, Paris, 1982.

[30] P. Hatziantoniou, D. Tsoukalas, J. Mourjopoulos, and S. Salamouris, "Time-Frequency Mapping Based on Nonuniform Smoothed Spectral Representations, " in Proc. *IEEE ICASSP-99*, Phoenix, AZ, 1999.

[31] P. D. Hatziantoniou and J. N. Mourjopoulos, "Generalized Fractional-Octave Smoothing of Audio and Acoustic Responses," *J. Audio Eng. Soc.*, vol. 48, pp. 259–280, 2000.

[32] S. G. Norcross, A. G. Soulodre, M. C. Lavoie "Subjective investigations of inverse filtering," *J.Audio Eng. Soc.*, vol. 52, pp. 1003-1028, 2004.

[33] H. Møller, D. Hammershøi, C. Jensen, and M. Sørensen, "Transfer characteristics of headphones measured on human ears", *J. Audio Eng. Soc.*, vol. 43, pp. 203–217, 1995.

[34] A. Farina, "Simultaneous measurement of impulse response and distortion with a swept-sine technique", *in Proc. 108th Audio Engineering Society Convention*, Paris, France, 2000.

[35] Sensory analysis – General guidance for the selection, training and monitoring of assessors – Part 2: Experts, ISO Standard 8586–2-2008.

[36] Methods for manual pure-tone threshold audiometry, ANSI Standard S3.21 - 2004, reaffirmed in 2009.

[37] R. Bücklein, "The audibility of frequency response irregularities," *J. Audio Eng. Soc.*, vol. 29, pp. 126–131, 1962, reprinted in English in 1981.

[38] S. Olive, P. Schuck, J. Ryan, S. Sally, and M. Bonneville, "The detection thresholds of resonances at low frequencies", *J. Audio Eng. Soc.*, vol. 45, 116–128, 1997.

[39] B.R. Glasberg and B. C. J. Moore, "A model of loudness applicable to time- varying sounds," *J. Audio Eng. Soc*, vol.50, pp 331-342, 2002.

[40] P. Rao, R. van Dinther, R. Veldhuis and A. Kohlrausch, "A measure for predicting audibility discrimination thresholds for spectral envelope distortions in vowel sounds," J. Acoust. Soc. Am., vol. 109, pp. 2085-97, May, 2001

[41] K. Okabe, H. Hamada and T. Miura, "Headphone responses on real ears and a head and torso simulator, " in *Proc. Acoustics, Speech, and Signal Processing, IEEE International Conference on ICASSP '86*, vol. 11, pp. 937-940, 1986.