Spatial Audio with Consumer Headphones: How its quality affects the immersion

P. Gutiérrez-Parera and J. J. López

Instituto de Telecomunicaciones y Aplicaciones Multimedia,
Universitat Politècnica de València,
8G Building - access D - Camino de Vera s/n - 46022 Valencia (Spain)
Corresponding author: pabgupa@iteam.upv.es

Abstract

A realistic sound immersion can be reproduced by high quality headphones. However, low quality consumer headphones are widely employed by common users. A weak frequency response, the distortion and the sensitivity disparity between the left and right transducers could be some of the degrading factors. In this work, we are studying how these factors affect spatial perception. A series or perceptual tests have been carried out with a virtual headphone listening test methodology. The first experiment focuses on the analysis of the disparity of sensitivity between the two transducers. The second test studies the influence of the frequency response relating perceived quality and spatial impression. The third test analyses the effects of distortion using a Volterra kernels scheme for the simulation of the distortion using convolutions. Finally, the fourth relate the quality of the frequency response with the accuracy on azimuth localization. The conclusions of the experiments are: the disparity between both transducers can affect the localization of the source; the perception of quality and spatial impression has a high correlation; the distortion produced by the range of headphones tested at a fixed level is not perceptible; and that some frequency bands have an important role in the front-back confusions.

Keywords: headphones; spatial sound; perceived quality; binaural; subjective test; distortion; frequency response; front-back confusion.

1. Introduction

Spatial audio technologies have gained great popularity in recent years, with the arrival of high definition TV, 3D video and mobile devices. Headphone-based systems have also grown in popularity in the last years, because of the private hearing they provide in any type of environment as well as the widespread use of smartphones and mobile devices. Headphones are commonly employed to reproduce stereo material, but binaural recordings increase the spatial hearing dramatically.

Binaural sound uses the principles of human auditory system [1] to reproduce the recordings over headphones. It assumes that, if we are able to reproduce in the listener's ears with headphones the same pressures that the listener experiences in a natural environment, a realistic acoustic immersion can be simulated [2].

To have a correct sense of spatial immersion, high quality microphones should be employed in conjunction with acoustic mannequins. In addition, high quality headphones should be used for playback. However, low-end headphones are widely used in most cases, either for economic reasons or simply because they are included with mobile devices. It is generally known that low cost headphones usually provide a poorer sense of immersion, but the degrading factors that cause such a loss in quality and perceived spatiality has not been sufficiently studied, as well as the level of their effects.

Different factors can affect the perception of the spatial sound image. Our hypothesis states that three main factors are responsible for this degradation. They could be the frequency response, the distortion and the disparity between the left-right transducers, especially in low cost headphones. To determine this, we propose a series of perceptual tests [3] to particularly study these factors.

Several previous works have studied headphones listening, usually attending to perceived quality [4] [5] and listeners' preference [6] [7] [8]. In this investigation we

focus on the influence of the headphones in the perception of spatial sound image.

Section 2 describes the methodology, the headphones employed in the study, as well as the technique used to measure and simulate them. Sections 3-6 explain a series of perceptual tests that constitute the bulk of this research. Firstly, Section 3 presents a perceptual test carried out to study the influence of the sensitivity disparity between left and right transducers and to establish how the perception of the sound source position in the azimuth is affected. Although in high quality headphones, manufacturers match transducers with similar sensibilities, these low cost headphones have different sensibilities due to broader manufacturing tolerances. Another second subjective perceptual test described in Section 4 was conducted to evaluate the effect of the frequency response in the perception of quality and spatial impression with headphones. As frequency response is the factor that varies most among different headphones due to their quality, this test is of particular interest to better understand how frequency response affects the spatial sound impression. Section 5 outlines the third perceptual test planned to evaluate the effect of harmonic distortion in listening with headphones. Distortion can be considerable if high dynamic sound and high reproduction levels are employed. Section 6 explains the fourth and last test, which studies the relation of the frequency response with the accuracy of localization in the horizontal plane. The capacity of a headphone to generate a good spatial immersion can be different from its capacity to generate precise locations. To explore this point, azimuth localization is tested here for different kinds of headphones. The discussion and conclusions of these experiments are presented in Section 7.

2. Methods

It is well known in loudspeaker testing that visual cues play an undesirable role in the results provided by test subjects. Similarly, when testing headphones tactile cues can also influence results. Consequently it can be challenging to conduct a double blind comparative listening test for headphones. It is difficult to hide the possible influencing variables such as brand, design or price. In addition, the manual substitution of different headphones on the subject's head can be disruptive and introduce useless fatigue on the subject [9]. Moreover, the fitting and tactile sensations are impossible to remove, making them an important bias factor [4].

In order to avoid these effects, it is appropriate to use a virtual headphone simulation to perform the listening tests [6] [10]. This method employs one reference headphone to simulate the different headphones under test. In this way, listeners can evaluate the simulated versions of the different headphones wearing just the reference headphone, therefore avoiding the manual change of headphones and removing the visual and tactile biases. Some other advantages are obtained with this virtual method: listeners can have immediate access to the different headphones and the procedure test become more flexible, transparent, controlled and repeatable.

The reliability of this virtual simulation method has been previously studied, finding good correlation between standard listening tests using real headphones and the virtual simulation method [11]. However, in some cases some discrepancy related to a specific model or sound signal [6] has been found due to the visual and tactile bias present in the standard test.

Because of the great advantages of a virtual test over a standard one, this study used a virtual headphone listening test methodology. This would remove the strong bias than would appear in this study due to the remarkable differences in appearance and fitting characteristics among the consumer headphones and the different range of qualities desired for this work.

2.1. Headphones selection

Different headphones were selected in order to represent a range of commercial and readily available headphones. According to this principle and the scope of the study described in previous sections, seven different headphones were selected plus a high quality reference one. A Sennheiser HD800 was chosen as the reference headphone (REF). The reason for this selection is due to its

Number	Abbreviation	Name	Type
1	REF	Reference, Sennheiser HD800	open and circumaural
2	HQop	High Quality open	circumaural
3	MQcl	Medium Quality closed	circumaural
4	BDso	Big Diaphragm semi-open	circumaural
5	LCmul	Low Cost multimedia	supra-aural
6	AirL	Airline	supra-concha
7	Wop	Wireless open	circumaural
8	LCmul2	Low Cost multimedia 2	supra-aural

Table 1. Headphones used in the study.

great fidelity, response, low distortion and accurate timbral reproduction. The other seven headphones were selected to cover a wide range of possible common use. The brands and models of the rest of the headphones will be omitted, as they are not necessary for the result analysis. The headphones classification used in the study is listed in Table 1.

The reference headphone was the only one that participants used, saw and had contact with during the tests. The rest of the headphones were simulated through the reference one.

2.2. Frequency responses measures

To measure the response of the different headphones, a swept-sine method was employed [12] using a Head and Torso Simulator (HATS) model B&K Type 4100 (Figure 1). This technique gave us the impulse response needed for the simulation of the different headphones.

Figure 1. Set-up for measuring the headphones with the Head and Torso Simulator (HATS).

To avoid differences in the amplitude level of the measures, the selected criterion was to achieve the same equivalent power between 100Hz to 10kHz for all the headphones (in order to minimize the influence of roll-off in low and high quality headphones). This decision allowed us to measure all the headphones in the same reproduction conditions and to achieve the same level in this band of frequencies. The reproduced pressure level for all the headphones was the equivalent to 69 dBSPL of pink noise in the reference headphones. This level was selected in informal tests as a pleasant listening level. Besides, this level allowed the measurement of the different headphones models without any saturation distortion in equivalent conditions.

Each of the headphones including the reference one, were measured with the mentioned swept-sine method. The resulting impulse responses ($h_n[n]$) were truncated to 50ms (2205 samples for 44100Hz sampling frequency) and windowed with a half Hamming window. This length provides good resolution in low frequencies until 20Hz. To minimize errors related to headphone positioning on the ear of the HATS simulator, five re-sets of the headphones were done and measured. The curves showed in Figure 2 are based on the average of those measures.

To avoid visual, tactile cues and other influencing variables (brand, design, price), a virtual headphone simulation was used to compare the different headphones in the listening tests.

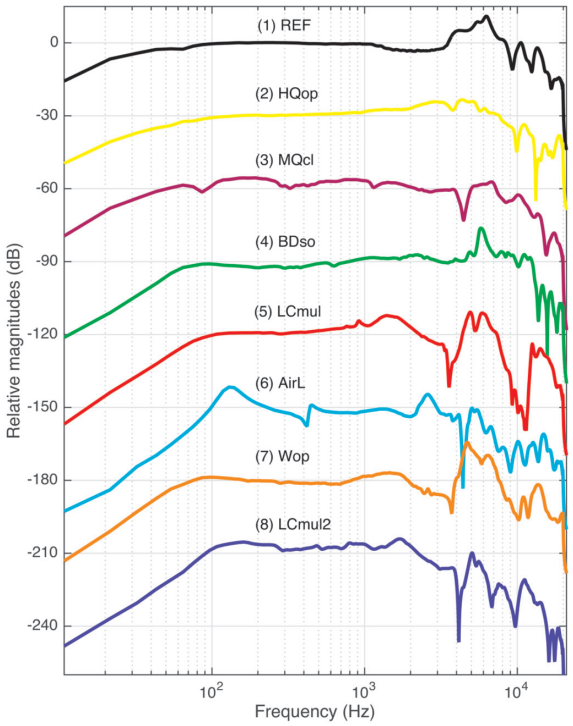


Figure 2. Frequency responses of the headphones used in the study (left channels, 30dB offset).

The first curve corresponds to the reference headphone (1)-REF which shows a smooth response and flat below 3kHz. The next three (2)-HQop, (3)-MQcl, (4)-BDso headphones were chosen as good-mid quality range with different characteristics: open, closed and semi- open. Their frequency responses below 6kHz are quite flat, with the exception of some irregularities in the (3)-MQcl curve and a peak down at 4.5kHz that decreases to -14dB. There is another peak up in the curve (4)-BDso at 6kHz of 15dB. The next curves (5 to 8) represent the frequency responses of the multimedia (5)-LCmul, airline (6)-AirL, wireless (7)-Wop and another multimedia (8)-LCmul2 headphones, that were chosen to be an example of mid and poor quality headphones. Their frequency responses have important peaks and valleys that affect the sound. Curve (5)-LCmul has a reinforcement in frequencies around 1.5kHz and a big dip in 3.5kHz, and curve (6)-AirL has a strong peak in 140Hz as well as other distortions up to 4.5kHz. Curve (7)-Wop is flatter in the mid frequencies with a small reinforcement in 1.5kHz and a

decay around 4.5kHz. In the case of curve (8)-LCmul2 it is important to note the rapid decline above 3kHz and the lack of proper high frequency beyond 5kHz. All these headphones are intended to be a small representation of quality range in commercial headphones.

2.3. Headphones frequency response simulation

The seven headphones under study were simulated to be reproduced with the reference headphones ((1)-REF-Sennheiser HD800). The simulation of each headphone was done filtering with its frequency response, but compensating the effect of the reference headphone using its inverted frequency response. Equation 1 shows the process for the simulation, where $H_n(\omega)$ is the measured response of the headphone to simulate, $H_{REF}(\omega)$ is the measured response of the reference headphone and $H_{n\ compensated}(\omega)$ is the response of the simulated headphone, which is applied to the corresponding stimulus.

$$H_{n\ compensated}(\omega) = \frac{H_n(\omega)}{H_{REF}(\omega)} \quad (1)$$

These virtual headphone equalizations include not only the magnitude response, but also the phase of the headphone measured. Although it is generally accepted that phase does not seem to affect the perceived accuracy of the simulations [13], especially if the stimuli material is typical music program, it can be noticed with pink noise stimuli. All the impulse responses of the headphones measured, the correction of the reference headphone and its application convolving with the stimulus, respect and keep the original phases. Moreover, accurate phase processing guaranties that our filtering will not alter in any way the Interaural Time Difference (ITD) between left and right transducers.

The filter implementation of Equation 1 was carried out in MATLAB in time domain, using Equation 2. Where $h_{n\ compensated}[n]$ is the response for the simulation of the virtual headphone, $h_n[n]$ is the impulse response of the headphone to simulate and $h^I_{REF}[n]$ is the inverted impulse response of the reference headphone.

$$h_{n\_compensated}[n] = h_n[n] * h^I_{REF}[n] \quad (2)$$

$h^I_{REF}[n]$ was calculated as follows: 1) the Fast Fourier Transform (FFT) of the h_{REF} was computed with a size of 4096 points with zero padding; 2) the resulting FFT was inverted; 3) to avoid an excess of boost at certain frequencies when correcting the reference headphone, the inverted response was limited to +15dB; 4) the Inverse FFT was properly computed and Hamming windowed to obtain the $h^I_{REF}[n]$. This process guaranties the avoidance of undesirable effects as circular convolution or others.

Finally, the different headphones were simulated applying the simulation filter $h_{n\ compensated}[n]$ to the sound materials for each test, obtaining the different stimuli.

2.4. Non-linear distortion simulation

As commented before, the swept-sine method employed to measure the frequency response of the headphones provides, besides the frequency response, distortion harmonics simultaneously. Figure 3 shows the frequency response and the second and third distortion harmonic of the reference ((a)-REF) and the airline ((b)-AirL) headphones. Both of these headphones are a good example of low (a) and high distortion (b).

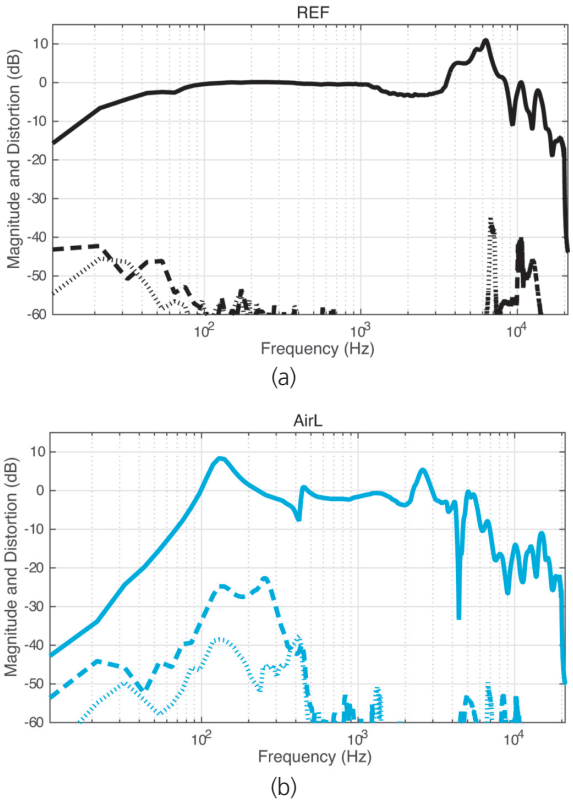


Figure 3. Frequency response with distortion of two headphones (left channel). Solid curve, magnitude; dashed and dotted curves, second and third order distortion harmonics. (a) REF headphone; (b) Airline headphone.

To simulate the non-linear distortion of each headphone, the method described in [14], which uses Volterra kernels and a series of linear convolutions, was chosen. With this method, the transfer function of a system is described by means of a Volterra series expansion. The output signal can be represented as the sum of the linear convolution of the measured impulse responses with the input signal and the corresponding frequency-shifted version. Applying Fourier transforms to these series results in a linear equation system. The solution of this system allows the computation of the diagonal Volterra kernels obtaining the impulse response terms for the main response and the first two distortion orders; Equation (3).

$$\begin{cases} H_1 = H'_1 + H'_3 \\ H_2 = -2\hat{H}'_2 \\ H_3 = -4H'_3 \end{cases} \quad (3)$$

where H_1', H_2', H_3' are the measured frequency responses and H_1, H_2, H_3 are the Volterra kernels (ˆ represents the Hilbert transform).

Using these equations, the second and third distortion orders were simulated by convolution, applying them to Equation (4), where $x(n)$ is the input signal and M is the number of samples of the kernel:

$$y(n) = \sum_{i=0}^{M-1} h_1(i) \cdot x(n-i) + \sum_{i=0}^{M-1} h_2(i) \cdot x^2(n-i) + \sum_{i=0}^{M-1} h_3(i) \cdot x^3(n-i) \quad (4)$$

More details of this technique can be found in [14]. This procedure was followed for Test 3.

2.5. Binaural Room Impulse Responses measures

In order to generate the spatiality of sound sources, some Binaural Room Impulse Responses (BRIR) were measured with a HATS B&K Type 4100.

Reverberation is an influential factor for spatial localization [1] [15] and because of this we decided to record our own BRIR with natural reverberation instead of using dry responses from a library. The impulse responses were recorded in a rectangular room with a volume of 132m^3 and a reverberation time of about 0.7s. Nine different azimuth angles were recorded (0º, 30º, 60º, 90º, 135º, 225º, 270º, 300º, 330º) in the horizontal plane at 1.5m of distance.

3. Test 1 – Sensitivity disparity between left-right transducers

3.1. Test description

The idea of this test is to evaluate how sensitivity disparity between the left and right transducers affects the perception of the source azimuth. To do that, a subjective perceptual test was carried out applying some volume level variations to different binaural sounds and checking how this affects the accuracy of horizontal localization.

In this test, participants had to listen, wearing headphones, to some binaural recordings obtained with a HATS on specific angles in the horizontal plane. Different variations of the original level between left and right transducers were applied to these sounds and then presented to the listeners. Participants should then indicate the direction of arrival, marking the angle in a Graphical User Interface (GUI).

The volume level variations applied were 0 (no modification), 1, 2 or 4 dB more on the left channel than the right one. Four different angles of direction of arrival were chosen, -30º, 0º, 65º and 90º of azimuth in the horizontal plane. Besides, the influence of different types of sounds was also studied.

These sounds were specifically recorded for this test using a binaural mannequin (B & K Model 4100) at the specific angles under study. A 44100 Hz sampling frequency was employed, obtaining full audio band recordings. The mannequin was in a semianecoic room, and sources were placed around it at 1m of dsitance. Four different sounds were recorded: a timbal drum hit, voice, a whistle and pink noise. The impulsivity of the timbal hit is an interesting characteristic regarding sound localization, also interesting for its low frequency content. Both voice and whistle are easily recognizable common sounds, which make them useful for the test. Moreover, the reduced spectral content of the whistle can be an interesting feature that can affect the test. The voice signal was the syllables "ba-be-bi-bo-bu", pronounced by a male voice. This sound has diverse vocalic contents and bilabial consonantal phoneme /b/, which produces impulsive sound. Pink noise was employed to evaluate a wide spectrum signal. All of these sounds were reproduced by the Sennheiser HD800 reference headphones.

According to the different types of sounds described above, the total number of stimuli presented to each participant in this test was: 4 angles × 4 types of sounds × 4 level variations = 64 stimuli. These stimuli were randomly presented, and the participant could listen to each of them as many times as he or she wanted.

During the test, participants also had the possibility of hearing a reference stimulus at any time, choosing between -90º, -45º, 0º, 45º and 90º of azimuth.

To perform the test, a simple GUI was developed in MATLAB that brings the user full control of the test. The participant could select the perceived sound source direction angle in an arc of -90º to 90º of azimuth (with a 5º resolution). It was also possible for the subject to freely control and listen to the reference stimulus (Figure 4).

Figure 4. Participant performing test 1.

The test was performed by 20 people, 10 men and 10 women (21 to 45 years, with an average age of 32). The average runtime of the test was 9 min. Every participant did a training session before taking the test, so all could listen to all of the stimuli and become familiar with the GUI and the assigned task. Some preliminary results of this test were previously published by the authors in [16].

3.2. Results

Figure 5a shows the average of the answered angles (for all of the level variation cases) according to the reproduced angle. The average of the answers has a deviation to the left-hand side. This is expected since the variations (0, 1, 2, 4 dB) were always more in favor of the left channel than the right.

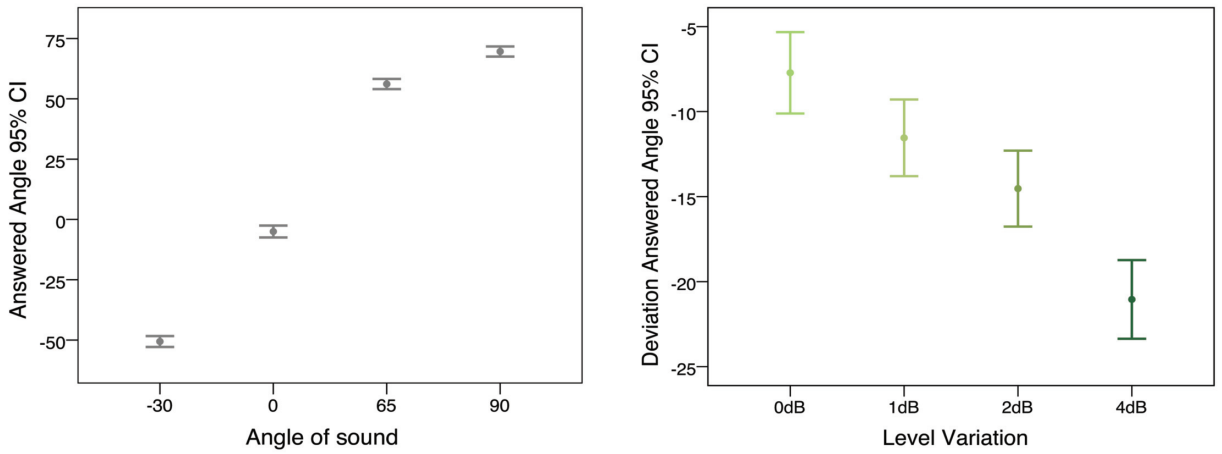


Figure 5. (a) Average of the answered angles versus reproduced angles (degrees); (b) average of the deviation of the answered angles (degrees) versus level variation (dB).

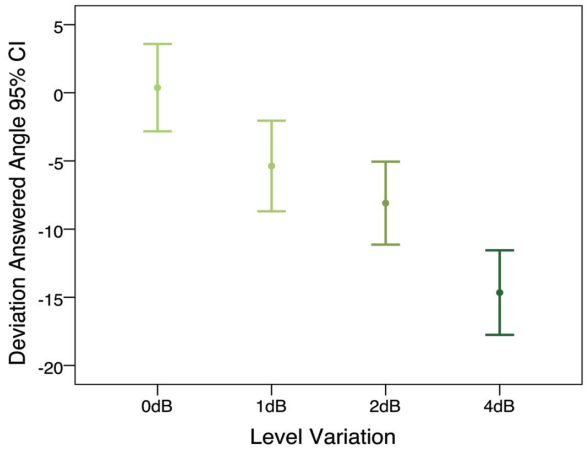


Figure 6. Average deviation of the answered angles (degrees) versus level variation (dB), considering only the angles 0° and 65°.

The tendency of this angle deviation to the left can be seen in Figure 5b, considering the level variation applied (0, 1, 2, 4 dB).

An Analysis of Variance (ANOVA) indicates that the level variation has a very significant influence ($F = 27.338$, df = 3, $p < 0.001$) over the deviation in the answers.

If we consider just the central angles used in the experiment (0° and 65°), a smaller average deviation can be seen (Figure 6). This leads us to believe that listeners tended to divert the location of the sounds perceived on the sides more, which means that the introduced level variations made the lateral angles disperse more than the central ones.

On the other hand, the influence of the type of sound (timbal, voice, whistle or pink noise) on the deviation in responses can be seen in Figure 7a. Voice and pink noise have lower deviation than timbal and whistle sounds, especially in cases of 0 and 1 dB of deviation. Besides, voice stimuli and pink noise manifest a more separate and clearer deviation at varying levels.

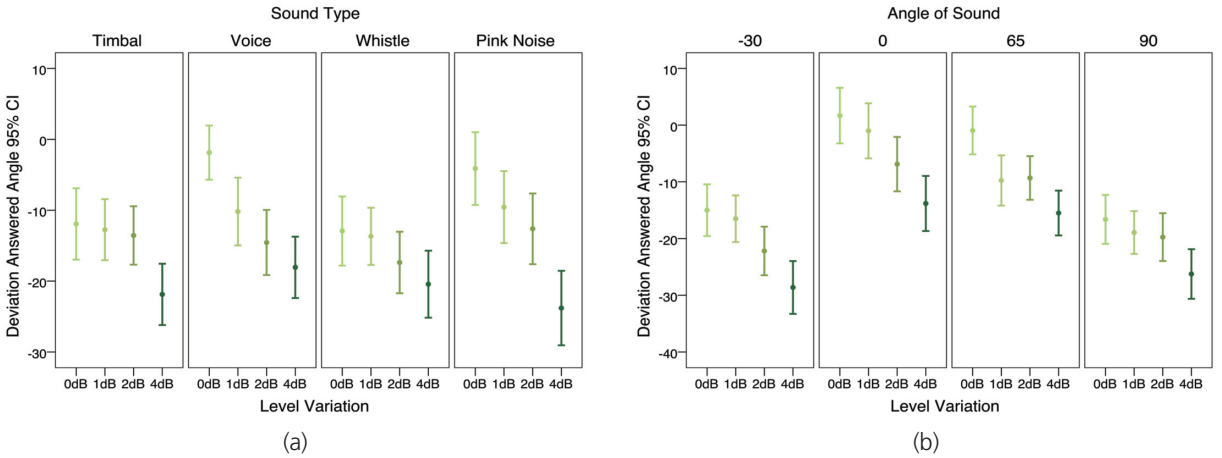


Figure 7. Average deviation of the answered angles (degrees) versus the level variation (dB): (a) considering the type of sound; (b) considering the angle reproduction of sound.

The influence of the type of sound over the deviation of answers is significant (F = 4.409, df = 3, p = 0.004) according to an analysis of variance. The sound angle reproduction has a very significant influence (F = 54.932, df = 3, $p < 0.001$) over the deviation of the answers. In Figure 7b, the deviation of the answers for each sound angle reproduction is represented. Angles 0º and 65º present less deviation to the left. The biggest deviation of the answers corresponds to the angle -30º, and it could be due to the fact that it was the only angle on the left side.

4. Test 2 – Frequency response about quality and spatial impression

4.1. Test description

In this test, participants listened to some excerpts of sound with headphones and rated their quality and their sound spatial image. These different headphones were simulated as described in Section 2.3 by means of the convolution of their frequency responses with the stimuli sounds, and all of them were reproduced with the reference headphones.

Due to the fact that different frequency responses produce noticeable effects, the perceptual test was designed according to the recommendation of the International Telecommunication Union, (ITU-R) 1534-2 [17], which describes the MUltiple Stimuli with Hidden Reference and Anchor (MUSHRA) perceptual test. This kind of test describes a method to assess intermediate quality audio systems and also all of the requirements needed to accomplish the test with rigor. Besides, this test sets a 0 to 100 continuous scale (0-bad; 100-excellent) to evaluate quality and other parameters of sounds and systems, always using a reference sound. All systems are compared to a reference of maximum quality, and the different systems are also compared between them.

Two different tasks were evaluated during the test by the participants. The first task was to indicate the quality of the sound with respect to the reference. The second task was to evaluate the spatial impression (locations, sensations of depth, immersion, reality of the audio event) [18] with respect to the reference.

Five different excerpts of audio (12 to 14s) were employed as source material (see Table 2), and all of them were reproduced simulating the different headphones under study. All of these sound fragments were chosen by their spatial, stereophonic and timbral attributes.

A high correlation was found between the subjective perception of quality of the headphones and the spatial impression.

In this test, five headphones simulations were done, corresponding to headphones (2)-HQop, (3)-MQcl, (4)-BDso, (5)-LCmul and (6)-AirL (described in Section 2.1, with frequency responses in Figure 2). Each of the five sound excerpts previously mentioned were reproduced by the virtual headphone simulation described in Section 2.3. A virtual headphone simulation for each sound was presented randomly in series to the listeners, as well as a hidden reference ((1)-REF) and also two anchor signals. The first Anchor signal (ANC1) was a 7-kHz low pass filtered version of the sound (according to the mid-quality anchor of the ITU recommendation 1534-2 [17]), and the second Anchor signal (ANC2) was a monaural version of the sound. This second anchor was determined to set a reference for the spatial impression question.

To perform the test, a GUI was developed in MATLAB according to the recommendation [17], which allowed participants to freely listen to each of the sounds and to the reference, as many times as they wanted (Figure 8). The different sound fragments were presented randomly as a series with all of the different headphone simulations, to compare to the reference sound. Once the participant had scored all of the simulations of a series, a new sound excerpt was presented to be evaluated. This process was repeated twice, once for each question of the test (the first about quality and the second about spatial impression), with a pause in between.

The number of stimuli of this test was: (5 headphones simulations + 1 hidden reference + 2 anchor signals) × 5 sound excerpts = 40 stimuli, presented in five series of eight stimuli plus the reference. As commented before, these 40 stimuli were presented twice in a different random order, to answer the two different questions.

The test was performed by 11 people, seven men and four women (21 to 37 years, with an average age of 30). As the test had two different questions, they were separated into two parts with a rest pause in the middle. The average runtime of the test was 22 min for the first part and 16 min for the second. Every participant did a training session before preforming the actual test, so all of them could listen to all of the stimuli and become familiar with the GUI and the assigned tasks.

Artist	Track	CD	Description
Bettina Flater	*Haugebonden*	Women en Mi	female voice and guitar
Paco de Lucía	*Zambra Gitana*	Canción Andaluza	male voice and guitar
Jerry Goldsmith	*Night Boarders*	OST The Mummy	high dynamic orchestral
The Chad Fisher Group	*Basin Street Blues*	live	jazz (binaural)
Smashing Pumpkins	audience sound	live	audience and drums (binaural)

Table 2. Music program used for listening tests 2 and 3.

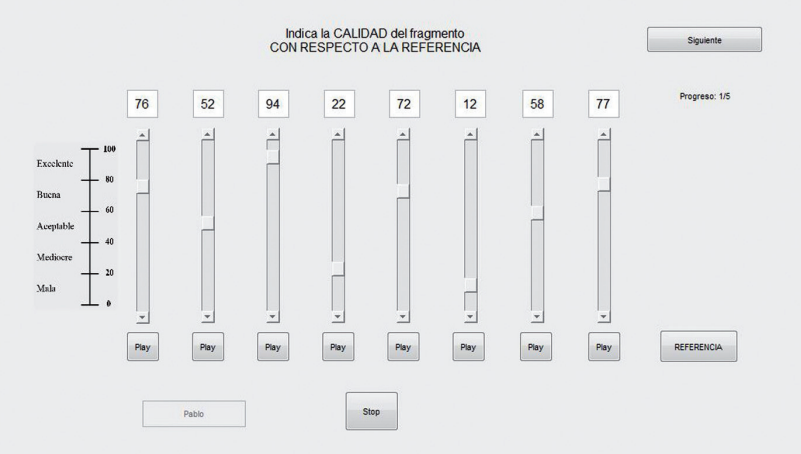


Figure 8. GUI of test 2.

4.2. Results

Figure 9a shows the average of the normalized (zero to 100) quality answers for the hidden reference, all five headphones simulated and the two anchors. As shown, the reference has been properly identified in most cases. The three supposedly good quality headphones have high scores; meanwhile, the two supposedly poor quality ones have the lowest scores. Both anchors remain in the middle of the scores of these two groups.

An analysis of variance confirms that the headphones have a very significant influence (F = 58.33, df = 7, p < 0.001) over the quality perceived.

Figure 9b shows the average of the normalized (0 to 100) spatial impression answers for the hidden reference, the five headphones simulated and the two anchors. The results seem to be similar to the answers about quality, with a high correlation of r^2= 0.648. Nevertheless, in this case, the confidence intervals are a bit wider, and the scores have some differences. The three supposedly good quality headphones have high scores again, but the confidence intervals do not separate them very much. There is a bigger difference between the two supposedly poor quality headphones, and the low cost multimedia ((5)-LCmul) ones are in the same range as both anchor signals. It is also noticeable that the Anchor Signal 2 (ANC2) as a monaural signal does not have a lower score.

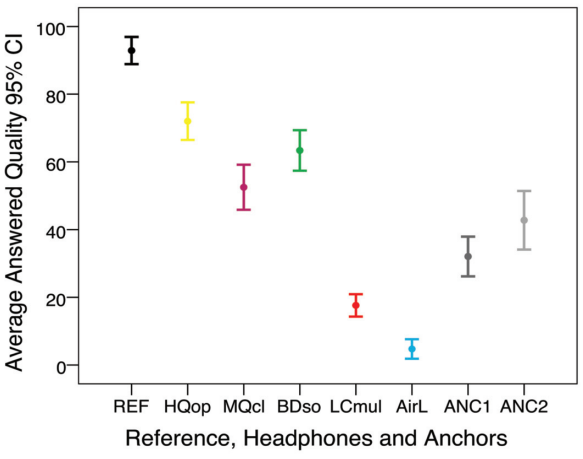
In any case, an ANOVA confirms that the headphones have a very significant influence (F = 58.33, df = 7, p < 0.001) over the perceived spatial impression. No significant influence of the type of sound has been detected, even though some of them were binaural recordings.

5. Test 3 – Non-linear distortion

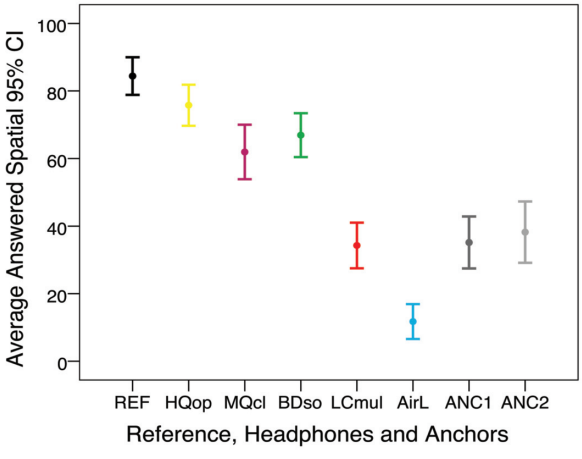
5.1. Test description

The objective of this test is to evaluate how the effect of harmonic distortion in headphones affects the spatial impression.

Several stimuli with and without the simulation of their harmonic distortion were presented to the participants that had to score their perception.



(a)



(b)

Figure 9. (a) Average answered quality versus reference, headphones and anchors; (b) Average answered spatial impression versus reference, headphones and anchors.

The effect of these distortions is very subtle. For that reason, the perceptual test was designed according to the recommendation ITU-R 1116-2 [19], which describes a method to assess small impairments in audio systems. This recommendation also establishes rigorous requirements of room, equipment and other arrangements. A continuous scale from one to five (1 - very annoying; 5 - imperceptible) is used to evaluate degradations with respect to a reference signal. The recommendation proposes an ABC test in which two stimuli, A and B, are presented to be compared against a known reference. One of these two stimuli, A or B, is always a hidden reference, and the other a degraded signal.

One single question was presented to the participants: "What degradation of quality and spatial impression do you hear with respect to the reference?"

The same five audio excerpts previously described in Test 2 were used here (see Table 2), as well as the same five virtual headphone simulations (2)-HQop, (3)-MQcl, (4)-BDso, (5)-LCmul and (6)-AirL (described in Section 2.1, with frequency responses in Figure 2). No anchors beyond the proposed scale were used this time.

Two different versions of the headphones simulations were presented in this test. One without and the other with the distortion simulated with the method described in Section 2.4. These two versions of the same stimulus were presented each time to the participants. They have then to rate the distorted against the not distorted version of the same sound in a double-blind manner (A *vs.* B). In each trial, there was always a non-distorted version sound that acted as the known reference (C sound), which according to the recommendation [19] has to be compared to the A and B sounds.

The number of stimuli of this test was then: 5 headphones simulations × 2 versions (with and without distortion) × 5 sound excerpts = 50 stimuli, presented in twenty five series of two stimuli plus the reference. All of these pairs were presented randomly to each participant.

To perform the test, a GUI was developed according to the recommendation, which allowed participants to freely listen to each of the sounds to evaluate and the reference, as many times as they wanted.

The five headphones under study were simulated (including distortion) to be reproduced with the reference headphones ((1)-REF, frequency response in Figure 2).

This test was performed by the same 11 people of the previous Test 2; seven men and four women (21 to 37 years, with an average age of 30). The average runtime of the test was 16 min. Every participant did a training session before preforming this test, so all of them could listen to all of the stimuli and become familiar with the GUI and the assigned task.

5.2. Results

According to the recommendation [19], the difference between the score of the hidden reference and the score of the degraded signal is analyzed. Figure 10 shows these differences for each of the headphones simulated.

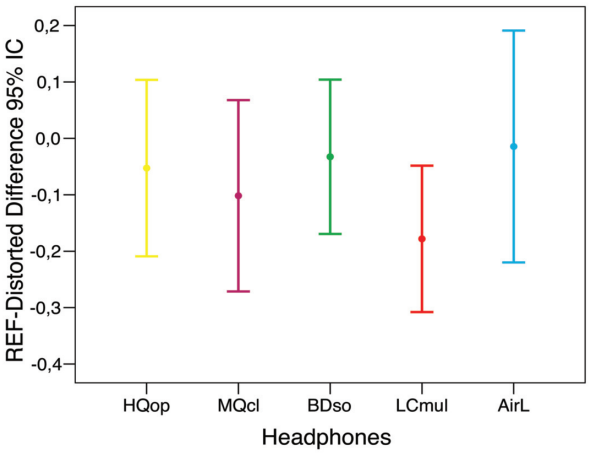


Figure 10. Difference between hidden reference and distorted signals versus headphones.

No significance has been found. Then, distortion can be considered as imperceptible. Therefore, it has no effect in spatial perception, at least with the fixed level used to simulate all headphones (69 dBSPL).

6. Test 4 - Frequency response about azimuth localization

6.1. Test description

The results obtained in Test 2 are significant, but do not provide information about the accuracy in the localization of sources. For that reason, a test to evaluate the influence of frequency response on this accuracy was carried out.

Attempts to describe different spatial attributes have been a constant pursuit in the field of spatial audio [18, 20, 21]. The diffuse term employed in Test 2 to ask about spatial characteristics (spatial impression) was intended to relate in a simple way the perception of quality with the feeling of spaciousness. A more specific study of spatial attributes is then necessary to better evaluate the performing of the different headphones. In this direction, the localization accuracy in azimuth is one of the most studied spatial attributes [22, 23, 24, 25] and therefore a good anchor point to contrast the previous Test 2 with a localization experiment. Therefore, this test tries to establish a relation of the influence of the frequency response on the azimuth localization in the horizontal plane.

As commented on in Section 2.5, to simulate the position of the sound sources in the horizontal plane, recordings of BRIRs in a medium sized room were done. Nine different azimuth angles, 0º, 30º, 60º, 90º, 135º, 225º, 270º, 300º and 330º were used.

Four types of sound were employed: door, voice (female), guitar and pink noise. A closing door is an impulsive sound with quite low frequency content, which can be useful for sound localization. The guitar sound was composed by various impulsive sounds in different main frequencies, one for each chord. Voice is an easily recognizable com-

mon sound, and female was chosen to have some energy in high frequencies. The words "*estímulo sonoro*" (*sound stimulus* in Spanish) were employed. They present the repeated fricative phoneme /s/ with high frequency content and the phoneme /t/, an occlusive articulation that generates impulsive sound. Pink noise was employed to evaluate a wide spectrum signal.

For this test, seven different headphones plus a hidden reference were simulated (Section 2.1). Besides these, an additional anchor auralization (low pass filtered (LPF) sounds at 7 kHz) for each angle was employed (ANC1).

Therefore, the number of stimuli in this test was: 9 angles × 4 types of sound × (7 headphones simulation + 1 hidden reference + 1 anchor auralization) = 324 stimuli. These stimuli were presented in random order in two parts of 162 stimuli, with a rest in between.

To perform the test, a GUI was developed in MATLAB (Figure 11), which allowed participants to freely listen to the stimuli from a random list as many times as they wanted. Participants should indicate the perceived angle of the sound source. The GUI consists of a circle of points, which represents the top view of the listener, with a 5º resolution. Additionally, it included a parallel control to freely listen to a reference sound (pink noise) in the angles of 0º, 45º, 90º, 135º, 180º, 225º, 270º and 315º.

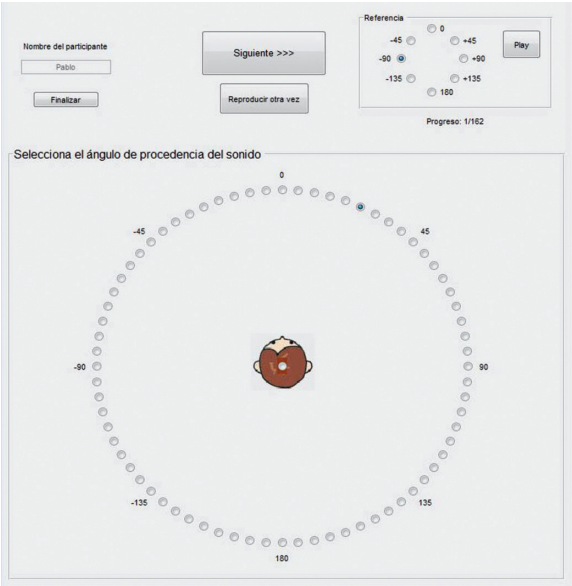


Figure 11. GUI of test 4.

The test was performed by 16 people, 10 men and 6 women (21 to 36 years, average age of 30). The average runtime was of 21 and 17 min for each part.

6.2. Results

A Cronbach's alpha analysis over the answers has been performed giving a value of $\alpha = 0.982$, which shows a high internal consistency.

A one-way ANOVA showed a significant influence between the headphones and the deviation of the answered angle (deviation = answered angle–real angle) ($F = 2.399$; df = 8; $p = 0.014$).

A first exploration of the participants' answers reveals that several front-back confusions [26, 27] occur. For this reason, an evaluation of the amount of front-back confusions was performed for each of the headphones simulated. An ANOVA showed that there is a very significant influence of the type of headphones on the number of front-back confusions ($F = 46.307$; df = 8; $p < 0.001$). In Figure 12, we can see that headphones (6)-AirL and (8)-LCmul2 produce an average of nearly 50% of front-back confusions. This can be logical, as both headphones are supposed to be in the low quality range. However, the (3)-MQcl headphone stands out in the group of high quality ones, as it has 30.2% of front-back confusions, more confusions than the (5)-LCmul headphone, with a significant difference. A comparison of the frequency response of the headphones that produce more front-back confusions ((6)-AirL, (8)-LCmul2 and (3)-MQcl) reveals that they share in common strong irregularities in the band of 100 to 1600 Hz. On the other side, other headphones of medium and low quality ranges that have less front-back confusions do not present these strong irregularities in that four-octave band. Because of that, we suspect this can be an affecting factor disturbing the front-back discrimination.

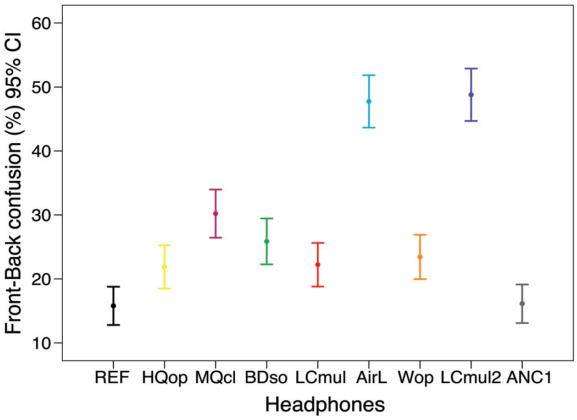


Figure 12. Percentage of front-back confusions for the reference, headphones and the anchor.

There is no significant influence of the type of sound crossed with the headphones. The guitar sound is the only one that produces slightly less front-back confusions for all of the headphones.

Due to the strong front-back confusion, the analysis of the deviation of the perceived sound with respect to the reproduced sound will produce large angle errors with complicated analysis of the results. A front-back confusion produces a bigger error for sources in the median plane than lateral sources, avoiding an analysis of the deviation angle (perceived angle – reproduced angle) with respect to the source position.

To overcome this setback, we propose a modified analysis of the error consisting of a preprocessing of the listener responses based on reflecting to the correct semi-plane

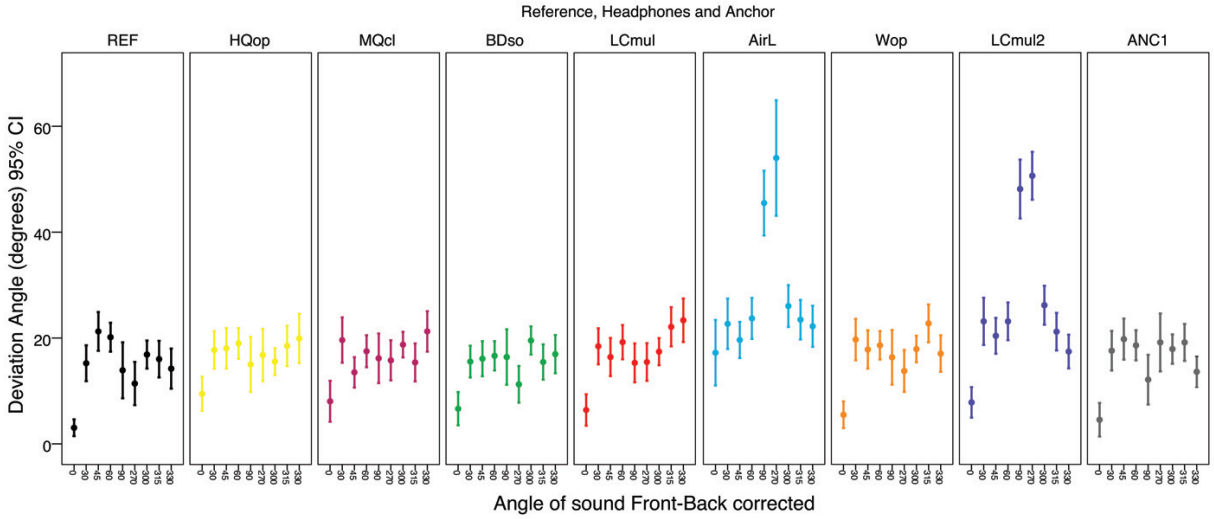


Figure 13. Deviation in degrees of the answers for each angle of sound reproduced. The reference, headphones under testing and anchor are represented.

the ones that have front-back confusion, leaving untouched the ones that do not. This correction eliminates big jumps in the deviation, focusing the experiment in the performance analysis of the headphones reproducing correctly the main spatial cues as ITD and the low frequency part of Interaural Level Difference (ILD). The high frequency part is more related to the pinna effect that is not considered with the reflection applied.

Taking into account the strong front-back confusion, the analysis of the answer deviation from the reproduction angle of the sound was performed introducing the correction of the front-back confusion. Therefore, a symmetric image of the responses in the back (90º to 270º) is brought to the front.

Figure 13 shows the deviation angle of the answers for the reproduction angle of the sounds, both of them front-back corrected. We can see that the deviations are quite uniform across the different headphones, except for the angles 90º and 270º in the cases of (6)-AirL and (8)-LCmul2. Looking at Figure 2, it is easy to see that the frequency responses of these two headphones present irregularities and deep level drops between 4 and 7 kHz. It is noticeable that the anchor LPF 7 kHz sounds auralized in the different angles (ANC1) are not affected by this problem, supporting the suspicion that the commented band is important for sources located in lateral positions.

7. Conclusions

This study outlines the influence of different quality parameters in headphones in the context of spatial sound reproduction. Four different perceptual tests have been done to analyze: (1) the effects of the sensitivity disparity between the transducers; (2) the influence of the frequency response over the perception of quality and the spatial impression; (3) the effects of non-linear distortion; and (4) the influence of the frequency response over azimuth localization.

The following main conclusions can be drawn:

1. The sensitivity disparities between left and right transducers affect the localization of sound sources, starting from level differences of 1 dB.

2. The quality and uniformity of the frequency response have an important influence in the *spatial impression*.

3. Additionally, the *spatial impression* has a high correlation with the subjective *perceived quality*.

4. The binaural recordings do not obtain significant better results for the parameter *spatial impression* compared to two-channel stereo mixes.

5. The distortion introduced by consumer level low quality headphones does not affect the perception of the spatial sound image.

6. It has been ratified that much front-back confusion is produced, both for high and low quality headphones.

7. We found that irregularities of the frequency response in the band of 100 to 1600 Hz seem to especially affect the front-back discrimination.

8. We also found that a poor response in the band of 4 to 7 kHz degrades the accuracy in lateral position localization.

All of these conclusions have been supported with statistical and ANOVA analysis. Some other interesting comments and clarifications about these conclusions can be added:

In addition to Conclusion 1, the angles chosen in the disparity test are a determining factor, whereby the more lateralized the angle, the larger the deviation. An increased number of angular positions may be of interest in later studies.

In relation to Conclusions 2 and 3, it is worth remarking that the mono anchor signal (ANC2) has obtained equal

or even better results for *spatial impression* than some headphones ((5)-LCmul, (6)-AirL) and the stereo LPF anchor (ANC1). This fact seems to be in relation to a deficient high frequency reproduction and the general listening sensation, as evidenced by the high correlation statistics obtained with the parameter *perceived quality*.

In relation to Conclusion 5, other works, such as [28], have not found significant perception of the distortion. However, this earlier study used high quality headphones, while ours does so also with low quality consumer headphones, and we have also analyzed the influence on spatial reproduction.

Finally, taking into account these three characteristics, *perceived quality*, *spatial impression* and accuracy in *azimuth localization*, we have concluded that the first two are highly correlated. Surprisingly, and contrary to how it might seem *a priori*, there is virtually no correlation between spatial impression and accuracy in localization, because the strong influence that the subjective perceived quality has over the spatial image perception. An illustrating example can be seen with the (5)-LCmul headphone. It would be interesting to deepen this relationship in future work.

Based on the results of this study, some general guidelines for the design of headphones suitable for spatial sound reproduction can be suggested. A sensitivity difference between left-right transducers less than 1 dB should be assured in the manufacturing process to avoid azimuth localization errors. A flat frequency response between 100 and 1600 Hz is desirable to reduce front-back confusion. Finally, a good frequency response in the band 4 to 7 kHz would guarantee a good accuracy in the localization of lateral sources.

Acknowledgments

The Spanish Ministry of Economy and Competitiveness supported this work under the projects TEC2012-37945-C02-01, TEC2015-68076-R and the grant BES-2013-065034.

References

[1] Blauert, J., Spatial Hearing: the psychophysics of human sound localization, MIT Press, Cambridge, Massachusetts, USA, 1997.

[2] Begault, D. R., 3D Sound for Virtual Reality and Multimedia Applications, Academic Press Professional Inc., San Diego, California, USA, 1994.

[3] Bech, S.; Zacharov, N. Perceptual Audio Evaluation - Theory, Method and Application; John Wiley & Sons Ltd.: Sussex, UK, 2006.

[4] Opitz, M., "Headphones Listening Tests," in *121st AES Convention, San Francisco, CA, USA*, 5-8 October 2006.

[5] Sung, H. Y., Kim, J., and Jang, S., "A Method for Objective Sound Quality Evaluations of Headphones," in *AES 32nd International Conference, Hillerød, Denmark*, 21-23 September 2007.

[6] Hirvonen, T., Vaalgamaa, M., Backman, J., and Karjalainen, M., "Listening Test Methodology for Headphone Evaluation," in *114th AES Convention, Amsterdam, The Netherlands*, 22-25 March 2003.

[7] Lorho, G., "Subjective Evaluation of Headphone Target Frequency Responses," in *126th AES Convention, Munich, Germany*, 7-10 May, 2009.

[8] Olive, S. E., Welti, T., and McMullin, E., "The Influence of Listeners' Experience, Age, and Culture on Headphone Sound Quality Preferences," in *137th AES Convention, Los Angeles, USA*, 9-12 October, 2014.

[9] Olive, S. E. and Welti, T., "The Relationship between Perception and Measurement of Headphone Sound Quality," in *133rd AES Convention, San Francisco, CA, USA*, 26-29 October, 2012.

[10] Briolle, F. and Voinier, T., "Transfer Function and Subjective Quality of Headphones: Part 2, Subjective Quality Evaluations," in *11th AES International Conference, Portland, Oregon, USA*, 29-31 May 1992.

[11] Olive, S. E., Welti, T., and McMullin, E., "A Virtual Headphone Listening Test Methodology," in *51st AES International Conference, Helsinki, Fin- land*, 22-24 August 2013.

[12] Farina, A., "Simultaneous measurement of impulse response and distortion with a swept-sine technique," in *108th AES Convention, Paris, France*, 18-22 February 2000.

[13] Lindau, A. and Brinkmann, F., "Perceptual Evaluation of Headphone Compensation in Binaural Synthesis Based on Non-Individual Recordings," *Journal of Audio Engineering Society*, 60(1/2), pp. 54–62, 2012.

[14] Farina, A.; Armelloni, E. "Emulation of not-linear, time-variant device by the convolution technique." In Proc. of the Congresso AES Italy 2005, Como, Italy, 3–5 November 2005.

[15] Rumsey, F., Spatial Audio, Focal Press, Oxford, UK, 2001.

[16] Gutierrez-Parera, P.; Lopez, J.J.; Aguilera, E. "On the influence of headphones quality in the spatial immersion produced by binaural recordings." In Proc. of the 138th AES Convention, Warsaw, Poland, 7–10 May 2015.

[17] Rec. ITU-R BS. 1534-2. Method for the Subjective Assessment of Intermediate Quality Level of Audio Systems; International Telecommunication Union (ITU): Geneva, Switzerland, 2014.

[18] Rumsey, F. "Spatial quality evaluation for reproduced sound: Terminology, meaning and a scene-based paradigm." *J. Audio Eng. Soc.* 2002, *50*, 651–666.

[19] Rec. ITU-R BS. 1116-2. Methods for the Subjective Assessment of Small Impairments in Audio Systems; International Telecommunication Union (ITU): Geneva, Switzerland, 2014.

[20] Letowski, T. "Sound quality assessment: Cardinal concepts." In Proceedings of the 87th AES Convention, New York, NY, USA, 18–21 October 1989.

[21] Zacharov, N.; Koivuniemi, K. "Unraveling the perception of spatial sound reproduction." In Proceedings of the 19th AES International Conference, Bavaria, Germany, 21–24 June 2001.

[22] Shinn-Cunningham, B. "Learning reverberation: Considerations for spatial auditory displays." In Pro-

ceedings of the International Conference on Auditory Display (ICAD), Atlanta, GA, USA, 2–5 April 2000.
[23] Minnair, P.; Olesen, S.K.; Christensen, F.; Møller, H. "Localization with binaural recordings from artificial and human heads." *J. Audio Eng. Soc.* 2001, *49*, 323–336.
[24] Santala, O.; Pulkki, V. "Directional perception of distributed sound sources." *J. Acoust. Soc. Am.* 2011, *129*, 1522–1530.
[25] Mendoça, C.; Campos, G.; Dias, P.; Santos, J.A. "Learning auditory space: Generalization and long-term effects." *PLoS ONE* 2013, *8*.
[26] So, R.H.Y.; Ngan, B.; Horner, A.; Braasch, J.; Blauert, J.; Leung, K.L. "Toward orthogonal non-individualised head-related transfer functions for forward and backward directional sound: Cluster analysis and an experimental study." *Ergonomics* 2010, *53*, 767–781.
[27] Zhang, P.X.; Hartmann, W.M. "On the ability of human listeners to distinguish between front and back." *Hear. Res.* 2010, *260*, 30–46.
[28] Temme, S.; Olive, S.E.; Tatarunis, S.; Welti, T.; McMullin, E. "The correlation between distortion audibility and listener preference in headphones." In Proceedings of the 137th AES Convention, Los Angeles, CA, USA, 9–12 October 2014.

Biographies

Pablo Gutierrez-Parera was born in Córdoba, Spain in 1982. He received a telecommunications engineer degree in 2008 from the Universidad Politécnica de Madrid. In 2010 and 2013 he obtained a M.S. degree in digital postproduction and a European graduate in telecommunication systems, sound and image engineering, both from the Universitat Politècnica de Valencia. Currently, he is a PhD grant holder from the Spanish Ministry of Economy and Competitiveness under the FPI program and is pursuing his PhD degree in telecommunications at the Institute of Telecommunications and Multimedia Applications (iTEAM) working in the field of spatial audio.

Jose Javier Lopez was born in Valencia, Spain, in 1969. He received the telecommunications engineer degree and the Ph.D. degree, both from the Universitat Politècnica de València, Valencia, Spain, in 1992 and 1999, respectively. Since 1993, he has been involved in education and research at the Communications Department, Universitat Politècnica de València, where he is currently a Full Professor. His research activity is centered on digital audio processing in the areas of spatial audio, wave field synthesis, physical modeling of acoustic spaces, efficient filtering structures for loudspeaker correction, sound source separation, and development of multimedia software in real time. He has published more than 200 papers in international technical journals and at renowned conferences in the fields of audio and acoustics and has led more than 30 research projects. Dr. Lopez was workshop co-chair at the 118th Convention of the Audio Engineering Society in Barcelona and has been serving on the committee of the AES Spanish Section for nine years, currently as secretary of the Section. He is a full ASA member, AES member and IEEE senior member.