Buenos Aires – 5 to 9 September, 2016
Acoustics for the 21st Century…

PROCEEDINGS of the 22nd International Congress on Acoustics

Communication Acoustics: Paper ICA2016-833

Evaluation of the frequency response of consumer headphones and its influence towards a binaural reproduction with HRTF individualization

Jose J. Lopez[(a)], Pablo Gutierrez-Parera[(b)]

[(a)] Universidad Politécnica de Valencia, Spain, jjlopez@dcom.upv.es
[(b)] Universidad Politécnica de Valencia, Spain, pablogparera@gmail.com

Abstract

Spatial audio technologies based on binaural techniques have sparked a renewed interest in recent years thanks to the spread of headphones listening using smart mobile devices. However, the quality of the headphones provided with smartphones or music players can be a determinant factor in the resulting binaural spatial sound sensation. In the first part of this paper, a perceptual test has been carried out to investigate the effects of the headphones frequency response over the azimuth localization in the horizontal plane. A virtual simulation headphone technique was employed to test seven different high and low quality headphones. Results indicate that much front-back confusion is produced in all cases and that some specific frequency bands have an important role on it and also a poor response in some high frequency bands can affect the lateral positions localization. In the second part of the paper a technique based on modelling the frequency response of the headphones and the HRTF using a chain of peak filters is presented. This technique is suitable to be employed for the individualization of the HRTF and it will solve some of the localization problems commented using less equipment than other published individualization methods.

Keywords: headphones, binaural, HRTF, individualization

Evaluation of the frequency response of consumer headphones and its influence towards a binaural reproduction with HRTF individualization

1 Introduction

With the introduction of high definition TV, 3D video and smart mobile devices, spatial audio technologies have sparked a renewed interest in recent years. Headphone-based systems have especially grown in popularity, because of the private hearing they provide as well as the widespread use of smartphones. Headphones are commonly employed to reproduce stereo material, but binaural recordings increase the spatial sensation 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]. However, there are two main limiting factors causing a not perfect satisfying result.

The first factor is related with the quality of the headphones employed for playback. High quality headphones should be used, however, low end headphones are widely employed 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.

The second are the anthropometric differences between subjects. A binaural signal recorded or synthesized for the ears of one subject is not generally valid for other, mainly due to differences in head and pinna size and shape. The Head-Related Transfer Function (HRTF) models the acoustic path of a plane wave that arrives at a certain angle to the subject's ears. It allows synthesizing binaural signals for both ears from a mono source at certain position in the space. Despite universal HRTF are employed more or less satisfactorily for spatial audio systems, some localization problems arise. Main problems are reduction of accuracy in azimuthal localization, front/back reversals, poor perception of elevated sources and lack of externalization.

In this paper we study the influence of the headphones frequency response over the azimuth localization in the horizontal plane. A perceptual subjective test [3] has been carried out to evaluate the effects of some real consumer headphones over a group of listeners. In section 4 an efficient technique for modelling the high frequency response of the HRTF by means of peak filter is presented that will permit a future HRTF individualization in an easy way.

2 Methodology

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 introduces useless fatigue on the subject [4]. Moreover, the fitting and tactile sensations are impossible to remove, making them an important bias factor [5].

In order to avoid these effects, it is appropriate to use a virtual headphone simulation to perform the listening tests [6,7]. 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 apart from other advantages. The reliability of this virtual simulation method has been previously studied [8], finding good correlation between standard listening tests using real headphones and the virtual simulation.

Seven different commercial headphones were selected plus a high quality reference one. A Sennheiser HD800 was chosen as the reference headphone (REF) due to its great fidelity, response, low distortion and accurate timbral reproduction. The other seven headphones were selected to cover a wide range of possible common uses (brands and models omitted).

The headphones used in the study were:

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

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. Then, all of the participants performed the test using the same high quality reference headphone (REF, Sennheiser HD800). The resulting signals for the rest of the headphones were simulated by means of proper signal processing algorithms and heard through the reference headphone.

2.1 Frequency Responses Measures

To measure the response of the different headphones, a swept-sine method was employed [9] using a Head and Torso Simulator (HATS) B&K Type 4100 (Fig. 1). 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 of the headphones (for calibration, we employed band pass pink noise between 100Hz to 10kHz for avoiding roll-off). This decision allowed us to measure all of the headphones in the same reproduction conditions and to achieve the same reproduction level in this band of frequencies. The reproduced pressure level for all of the headphones was equivalent to 69 Sound Pressure Level dB (dBSPL) of pink noise in the reference headphones. This level was selected in informal tests as a pleasant listening level.

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

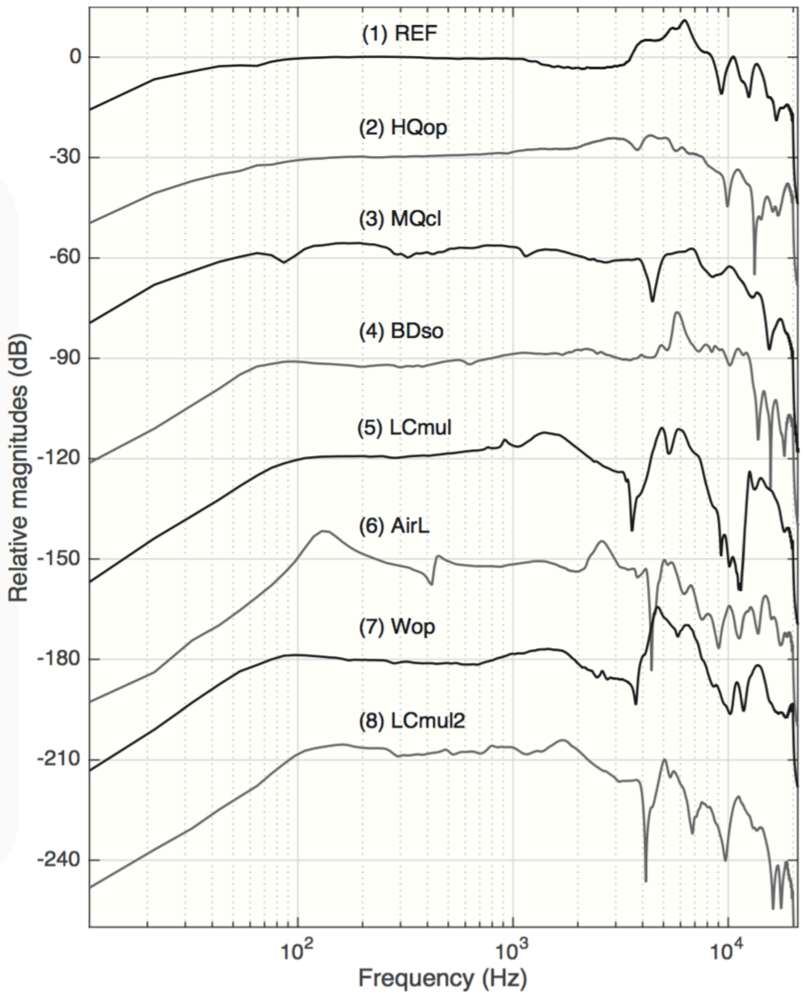

Figure 2: Frequency response of the headphones used in the study.

Each of the headphones, including the reference one, were measured with the mentioned swept-sine method. The resulting impulse responses (hi[n]) were truncated to 50ms (2205 samples for a 44,100-Hz sampling frequency) and windowed with a half Hamming window. This length provides good resolution in low frequencies until 20 Hz. To minimize errors related to headphone positioning on the ear of the HATS simulator, five resets of the headphones were done and measured. The curves shown in Fig. 2 are based on the average of those measures.

3 Experiments

3.1 Test description

Attempts to describe different spatial attributes have been a constant pursuit in the field of spatial audio [10-13]. The localization accuracy in azimuth is one of the most studied spatial attributes [14–17] and therefore a good anchor point to contrast with previous experiments. Therefore, this test tries to establish a relation of the influence of the frequency response on the azimuth localization in the horizontal plane.

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 common 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, 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 5◦ resolution. Additionally, it included a parallel control to freely listen to a reference sound (pink noise) in the angles of 0, 45, 90, 135, 180, 225, 270 and 315 degrees.

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.

3.2 Results

A Cronbach's alpha analysis over the answers has been performed giving a value of α = 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 [18,19] 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 3, we can see that headphones (f)-AirL and (h)-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 (c)-MQcl headphone stands out in the group of high quality ones, as it has 30.2% of front-back confusions, more confusions than the (e)-LCmul headphone, with a significant difference.

A comparison of the frequency response of the headphones that produce more front-back confusions ((f)-AirL, (h)-LCmul2 and (c)-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.

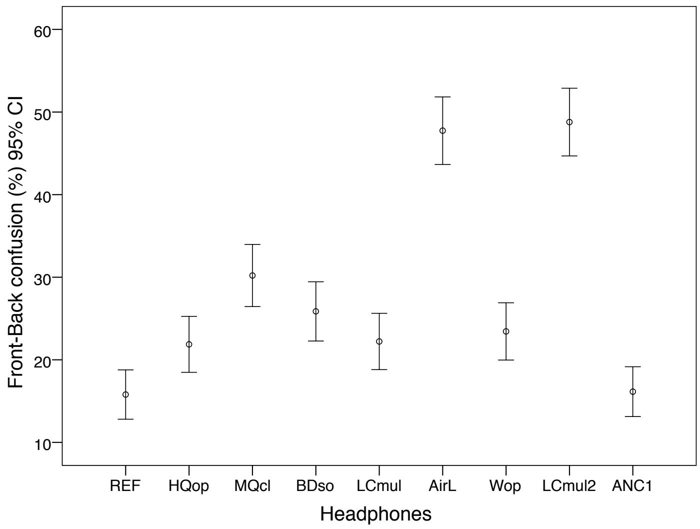

Figure 3: Percentage of front-back confusions for the reference, headphones and the anchor.

There is no significant influence of the type of sound crossed with the headphones. The sound guitar 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 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 Interaural Time Difference (ITD) and the low frequency part of Interaural Level Difference (ILD). The high frequency part that is more related to the pinna effect 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 4 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 (f)-AirL and (h)-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.

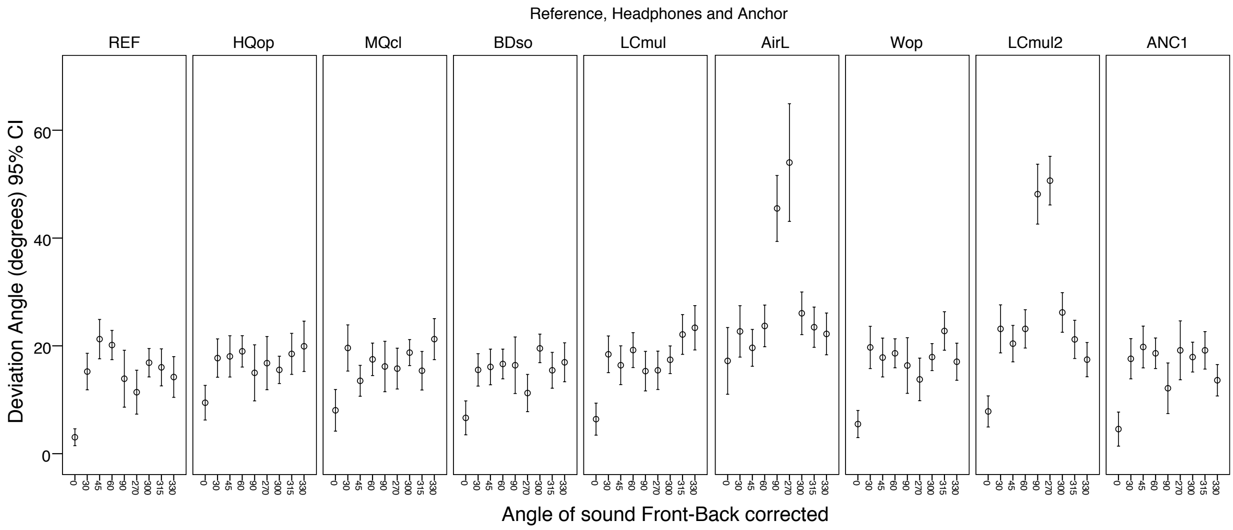

Figure 4: Deviation in degrees of the answers for every reproduced angle of sound. The reference, headphones under testing and anchor are represented.

4 Modelling HRTF for individualization

As it has been shown in previous experiments, the frequency response quality of headphones significantly affects location accuracy. The front-back confusion is the most noticeable effect, but errors in lateral sources location are also important, even with the high quality headphones.

Individualization of HRTF does not solve these problems completely but improves all of them until certain point. There are different methods of HRTF individualization previously explored in the literature as physical models based on anthropometric measures, statistical models, scaling models and off course to measure the HRTF of the subject with microphones in the ears. However, the previous methods require expensive hardware, are time consuming or require specialized technicians to carry out.

With the experienced gained testing headphones, we propose a method based on the correction of the frequency response in base to the position and amplitude of the different peaks and notches that can be identified in the frequency response between 5 kHz and 20 kHz. This method was previously employed by the authors in [20] for modelling elevation with good results.

For this purpose, a bank of parametric peak filters in series model the different oscillations of the frequency response, adjusting the parameters of the filters. A peak filter is a second-order parametric filter, which enhances or attenuates with a gain G (dB) at a bandwidth centered at a frequency fc, and with a quality factor Q. Equation (1) shows the normalized analog prototype peak filter where A is the square root of the filter gain in volts.

$$H_{peak}(s) = \frac{s^2 + s\frac{A}{Q} + 1}{s^2 + s\frac{1}{AQ} + 1} \tag{1}$$

This minimum phase filter will be transformed to the discrete domain H(z) by bilinear transformation, and the digital filter coefficients a1 , a2 , b0 , b1 and b2 from the filter parameters fc, Q, G are computed. Therefore, a high efficient structure using N peak filters in cascade can be used to model the modulus of the HRTF frequency response. At this point, the next problem to solve is the estimation of the 3 parameters fc, Q and G for each of the N filters and for any azimuth.

The parametrization of the HRTF for each azimuth by means of peak filters provides a powerful starting point for the adaptation to different individuals. Starting from a general HRTF it is possible to adapt it to each person modifying a small number of parameters corresponding to the mentioned peak filter parameter fc, Q, G and moreover, a listening/training method can be easily adapted for this purposes. Based on an easy psychoacoustic test and an interactive software the user, wearing headphones, sets up a personalized HRTF step by step. The algorithm starts with a generalized HRTF on which that small modifications are made in order to test if the spatial perception is improved or degraded. The person under test is asked in a clearly way using a graphic user interface where different spatialized sounds can be played. This technique can be initially focused on trying to model the pinna filtering effect in relation to from/back localization. This iterative algorithm will explore specific aspects of perception in a smart way in order to converge to a good HRTF based on the subject answers within a reasonable time.

As a future work, a subjective study will be carried out where front/back reversals using a universal HRTF and the ones obtained using our individualization method will be compared.

5 Conclusions

This study outlines the influence of the quality in headphones in the context of spatial sound reproduction. A perceptual test has been carried out to analyze the influence of the frequency response over azimuth localization.

The following main conclusions supported with statistical and ANOVA analysis can be drawn from the test:

1. It has been ratified that much front-back confusion is produced, both for high and low quality headphones.
2. We found that irregularities of the frequency response in the band of 100 to 1600 Hz seem to especially affect the front-back discrimination.
3. We also found that a poor response in the band of 4 to 7 kHz degrades the accuracy in lateral position localization.

In the second part of the paper an efficient technique for modelling the high frequency response of the HRTF has been presented. This technique is focused of trying to model the pinna filtering effect in relation to from/back localization. For this purpose, a bank of parametric peak filters located between 5kHz and 15 kHz will be optimized in order to reduce the front/back reversals. As a future work, a subjective study will be carried out where front/back reversals using a universal HRTF and the ones obtained using our individualization method will be compared. We expect this method will obtain similar or better improvement of other more complicated methods previously published but without using special equipment.

Acknowledgments

The Spanish Ministry of Economy and Competitiveness supported this work under the project TEC2015-68076-R and the grant BES-2013-065034.

References

[1] Blauert, J. Spatial Hearing: The Psychophysics of Human Sound Localization; MIT Press: Cambridge, MA, USA, 1997.

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

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

[4] Olive, S.E.; Welti, T. The relationship between perception and measurement of headphone sound quality.In Proceedings of the 133rd AES Convention, San Francisco, CA, USA, 26–29 October 2012.

[5] Opitz, M. Headphones listening tests. In Proceedings of the 121st AES Convention, San Francisco, CA, USA, 5–8 October 2006.

[6] Hirvonen, T.; Vaalgamaa, M.; Backman, J.; Karjalainen, M. Listening test methodology for headphone evaluation. In Proceedings of the 114th AES Convention, Amsterdam, The Netherlands, 22–25 March 2003.

[7] Briolle, F.; Voinier, T. Transfer function and subjective quality of headphones: Part 2, subjective quality evaluations. In Proceedings of the 11th AES International Conference, Portland, OR, USA, 29–31 May 1992.

[8] Olive, S.E.; Welti, T.; McMullin, E. A virtual headphone listening test methodology. In Proceedings of the 51st AES International Conference, Helsinki, Finland, 22–24 August 2013.

[9] Farina, A. Simultaneous measurement of impulse response and distortion with a swept-sine technique. In Proceedings of the 108th AES Convention, Paris, France, 18–22 February 2000.

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

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

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

[13] Zacharov, N.; Koivuniemi, K. Unravelling the perception of spatial sound reproduction. In Proceedings of the 19th AES International Conference, Bavaria, Germany, 21–24 June 2001, Appl. Sci. 2016, 6, 117 18 of 18

[14] Shinn-Cunningham, B. Learning reverberation: Considerations for spatial auditory displays. In Proceedings of the International Conference on Auditory Display (ICAD), Atlanta, GA, USA, 2–5 April 2000.

[15] 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.

[16] Santala, O.; Pulkki, V. Directional perception of distributed sound sources. J. Acoust. Soc. Am. 2011, 129, 1522–1530.

[17] Mendoça, C.; Campos, G.; Dias, P.; Santos, J.A. Learning auditory space: Generalization and long-term effects. PLoS ONE 2013, 8, e77900.

[18] 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.

[19] Zhang, P.X.; Hartmann, W.M. On the ability of human listeners to distinguish between front and back. Hear. Res. 2010, 260, 30–46.

[20] Lopez, J.J.; Cobos M.; Pueo B., Elevation in Wave-Field Synthesis Using HRTF Cues. Acta Acustica 2010, 96, 340-350.