Choosing Audio Devices on the Basis of Listeners' Spatial Perception:

A Case Study of Headphones vs In-earphones

Carlos Silva
Department of Informatics
University of Minho
Centro de Computação Gráfica
Guimarães, Portugal
Carlos.Silva@ccg.pt

Sandra Mouta & Jorge Santos
ALGORITMI Research Center
University of Minho
Centro de Computação Gráfica
Guimarães, Portugal

***Abstract*— The Earphones and Headphones industry is steadily growing following the emergence of new technological advancements and new applications. New methods to determine listeners' performance using different types of audio output devices will be in high demand. In this paper we adapt a methodology for evaluation of listeners' auditory localization accuracy to support the choice between two devices. As a case study, we compare a particular set of in-earphones and headphones. Our goal was to present a method that allowed us to: (1) conclude which audio device provided the most accurate sense of auditory localization; (2) understand the effect of training on task performance; and (3) determine which type of device benefits the most from short sessions of training in auditory localization. Participants had better performances using headphones. Nevertheless, we can reduce the differences between devices if short training sessions are included and the same device is used between training and test.**

Keywords— Consumer Electronics: Audiosystems–Headphones; ThreeDimensional Graphics and Realism—Virtual Reality; Physics: Acoustics—Psychoacoustics

I. Introduction

The recent growth of immersive technology in the consumer market is mainly due to new technological advancements in visual displays. Nonetheless, both researchers and manufacturers have already acknowledged that for successful immersive experiences, it is also important to create an appropriate and congruent immersive listening environment [1]-[3]. Thus, audio output devices, particularly wearable ones, will play a major role in the transition between commercial visual immersive systems to commercial audiovisual immersive systems [1].

The Earphones and Headphones industry has been steadily growing and follows the emergence of new technological advancements – as noise canceling and wireless technology – and new applications – like the incorporation of 3D sound in virtual reality systems. As applications requiring spatialized sound make their way into the market, new methods to determine listeners' performance will be in high demand. These assessments are of particular interest for immersive virtual environments (IVEs) developers that are looking for the best audio devices to support auditory stimulation. The integration of spatial sound in IVEs has been positively correlated with the feeling of presence [4] and the IVEs industry is already aware of the benefits that one can gather when more effort is focused on sound rendering (see, for instance, the collaboration between Oculus Rift and RealSpace^{TM} 3D audio).

The process of rendering audible, by physical or mathematical modeling, the sound field of a source in a virtual space is referred to as *Auralization* [5] [6]. The most widespread method for auralization and acoustic simulation takes into account the listener's anatomy – head, *pinnae,* and ear canal shape – and simulates its effect on the sound wave [6]. The listener's anatomy affects mainly the inter-aural time and inter-aural level differences (ITD and ILD respectively), which are the main static cues for sound location [7]. Thus, we can simulate a given position of the sound source in azimuth and elevation, by filtering an anechoic sound through a function that shapes each channel output giving it the accurate ITD and ILD for that position in space. These functions are called *Head Related Transfer Functions* (HRTFs). Auralization using HRTFs seems to be an appropriate solution for commercial applications, particularly the ones using databases of non-individualized HRTFs (captured using Head and Torso simulators). Studies have shown that listeners can locate non-individualized HRTF-based sounds [7] and that short training sessions improves significantly the localization performances [8].

In this study we present a method that allowed us to find out if listener's performance on auditory location tasks using non-individualized HRTFs is dependent on the type of audio devices used. This question is particularly interesting when we compare headphones and in-earphones, because the former devices allow individualized *pinnae* and ear-canal modulation over the non-individualized HRTFs, while the latter devices do not.

II. Methodology

A. Participants

16 participants with no previous experience in laboratory controlled auditory location tasks. All participants had normal

This work was supported by Bial Fundation Grant 143/14.

hearing, measured by standard audiometric tests. None showed inter-aural sensitivity differences above 5dB HL.

B. Conditions

Two conditions regarding audio output device (headphone VS in-earphone) in experimental phases (**intra-subject**); two groups of eight participants each regarding audio output device in training phase (**inter-subject**).

C. Material

Fig. 1. Audio output devices used in the experiment. Headphones – Sennheiser HD 650; In-earphones – Etymotic ER-4B Micro Pro.

D. Stimuli

A three second duration anechoic Pink Noise, auralized using HRTFs taken from the MIT database [9]. We present 18 different source positions in the horizontal plane (i.e., elevation 0º), with azimuth ranging from front to right in steps of 6º, from azimuth -6º to azimuth 96º. All sounds were auralized as free-field presented at 1 meter from the listener. Free-field means that only directional cues were presented and room acoustic cues were absent. The sound output intensity was measured and matched for both audio output devices, using a Brüel & Kjær type 4128C head and torso simulator and a PULSE™ acoustic analyzer platform.

E. Procedure

We adapted a procedure previously developed in our laboratory [7]. The overall experiment consisted of three phases:

(1) *Pre-training phase* where all stimuli were randomly presented (with four repetitions each) and after each stimulus presentation its localization was estimated in a touch-screen (see *Fig. 2*, panel A);

(2) *Training phase* where for five minutes participants could freely listening to five stimulus correctly positioned in the answer interface (see *Fig. 2*, panel B). At the end of this time participants listened each one of the five trained sounds and should click on the correct rectangle. Correct feedback was given at the end of each trial and this phase would end when participants reached an 80% correct answer level of performance;

(3) *Post-training phase*, where participants repeated the same procedure as in the pre-training phase.

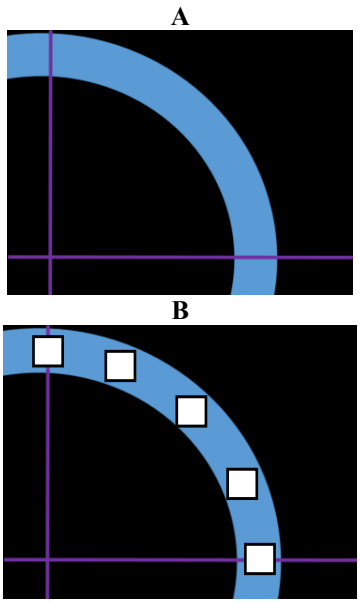

Fig. 2. Answer interface. **Panel A** – Participants were required to estimate the sound position in azimuth along the purple arch. **Panel B** – 5 positions of the trained stimuli. The answers were collected in a touchscreen, using a touchscreen stylus in order to increase precision.

III. Results

Table 1 shows the absolute mean error in degrees azimuth for each audio device used in each experimental phase.

TABLE I. PERFORMANCE BY CONDITION AND EXPERIMENTAL PHASE

N = 16	**Data grouped by device used in the Experimental phase**	
	Azimuth_Pre-training	*Azimuth_Post-training*
Headphones Abs Mean Error	16.77º (SD=4.79)	13.88º (SD=4.62)
In-ear Phones Abs Mean Error	18.83º (SD=6.74)	17.97º (SD=8.04)

The absolute mean error is lower on the Headphones condition, for both the Pre-training and the Post-training sessions. Paired sample t-test revealed significant differences between listening devices for the absolute mean error in the Post-training session ($t(15) = -2.513$, $p<.05$). A difference of 4.02º in the post-training results, corresponds to a sound displacement of approximately 7 cm, at 1 meter from the listener.

Fig. 3 presents the absolute mean errors distribution as a function of the stimuli position, for both conditions.

978-1-5090-2096-6/16/$31.00 ©2016 IEEE

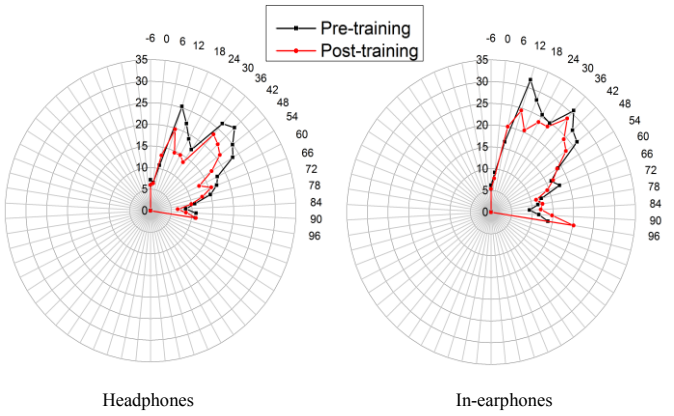

Fig. 3. Polar graphics with the absolute mean error as a function of the stimuli position.

As we can see from *Fig. 3*, the localization errors are higher in intermediate azimuths and lower on the ear plane and on frontal regions. This pattern of response is present with both equipments, however there are globally lower errors in the headphones condition and that is even more clearly observed in the extreme presentations (ear plane and frontal regions).

In a second analysis, we grouped the participants by audio output device used during training sessions. In doing this we wanted to understand how congruency regarding devices used on training and experimental sessions might affect performance on auditory location.

TABLE II. DATA GROUPED BY TRAINING LISTENING DEVICE

N = 8	**Data grouped by training listening device - *Headphones***	
	Azimuth_Pre-training	*Azimuth_Post-training*
Headphones Abs Mean Error	17.24º (SD=4.97)	13.84º (SD=5.24)
In-ear Phones Abs Mean Error	17.64º (SD=5.61)	20.13º (SD=10.33)
N = 8	**Data grouped by training listening device – *In-earphones***	
	Azimuth_Pre-training	*Azimuth_Post-training*
Headphones Abs Mean Error	16.36º (SD=4.94)	13.93º (SD=4.25)
In-ear Phones Abs Mean Error	19.85º (SD=7.97)	15.81º (SD=4.85)

From Table 2 we can see that keeping congruency (grey cells) between listening devices used during training and experimental phases, gives rise to generally lower absolute mean errors of sound localization in the post-training phase. Incongruency between training and experimental session listening device disrupted completely the benefits of training in the case of participants that used in-earphones in experimental phases. A mean decrement in performance of about 2.5º is observed for these participants, from pre to post-training session (also the mean value presents more variability). Nevertheless, incongurency did not prevent learning and better performance in post-training sessions for participants that used headphones in expeirmental phases.

Fig. 4 presents the distribution of the mean error as a function of the stimuli position, for the congruent sessions (same audio output device in training and experimental sessions). In *Fig. 4*, positive errors indicate misjudgments in sound location torwards the ear plane, while negative errors indicate misjudgments of sound location torwards the frontal plane (azimuth 0º).

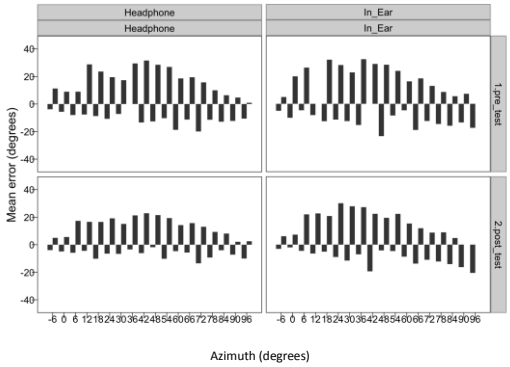

Fig. 4. Mean error distribution and direction as a function of the stimuli position, for the congruent sessions. Positive errors indicate misjudgments in sound location torwards the ear plane, negative errors indicate misjudgments of sound location torwards the frontal plane.

Interestingly it is possible to observe that positive errors are predominant, meaning that when misjudging location participants are prone to locate the stimulus as closer to the ear plane.

Finally, as headphones are more permeable to external noise when compared with in-earphones, we conducted a test to verify if the results obtained in silent conditions would hold in conditions with added environmental noise. Thus, we replicated this experimental protocol for eight new participants in a set-up in which the environmental noise reached the 56 dB(A) SPL. In these environmental conditions participants had an absolute mean error of 18.52º azimuth for the pre-training session, and an absolute mean error of 14.86º azimuth for the post-training session. These results differ on an average of 1.16º, when compared with results of congruent sessions using headphones.

IV. DISCUSSION

We presented a valuable method to access listener's spatial perception and evaluate performance between two audio devices. In the comparison between these particular models, headphones appeared to be the best solution for presentation of auralized sound and we should further investigate the benefits of using large housing with open back headphones. The fact that large housing headphones may allow individualized *pinnae* and ear-canal modulation over the non-

978-1-5090-2096-6/16/$31.00 ©2016 IEEE

individualized HRTFs, might be an important factor in the final performance outcome.

Nevertheless, we can reduce the differences between devices if short training sessions are included and the same audio output device is used between training and test. In-earphones can benefit greatly of maintaining congruency between experimental and training phases.

Future work should exhaustively compare between several types of audio output devices and should also investigate how performance is affect by the introduction of binaural room acoustic cues.

REFERENCES

[1] T. Lee, Y. Baek, Y.-C. Park, and D. H. Youn. "Stereo upmix-based binaural auralization for mobile devices." *IEEE Transactions on Consumer Electronics* 60.3 (2014): 411-419.

[2] S.-W. Jeon, Y.-C. Park, and D. H. Youn. "Acoustic depth rendering for 3D multimedia applications." *In Proc.* of the *IEEE International Conference on Consumer Electronics,* Las Vegas, USA, pp. 253-254, Jan. 2012.

[3] S. Kim, Y. W. Lee, and Y. J. Lee. "3D sound rendering system based on relationship between stereoscopic image and stereo sound for 3DTV." *In Proc. of the IEEE International Conference on Consumer Electronics,* Las Vegas, USA, pp.324-325, Jan. 2013.

[4] S. Poeschl, K. Wall, and N. Doering. "Integration of spatial sound in immersive virtual environments an experimental study on effects of spatial sound on presence." *2013 IEEE Virtual Reality (VR)*. IEEE, 2013

[5] J. E. Summers. "What exactly is meant by the term “auralization?”." *The Journal of the Acoustical Society of America* 124.2 (2008): 697-697.

[6] M. Kleiner, B.-I. Dalenbäck, and P. Svensson. "Auralization-an overview." *Journal of Audio Engeneering Society* . 41.11 (1993): 861-875.

[7] C. Mendonça, G. Campos, P. Dias, J. Vieira, J. P. Ferreira, and J. Santos. "On the improvement of localization accuracy with non-individualized HRTF-based sounds." *Journal of Audio Engeneering Society* 60.10 (2012): 821-830.

[8] C. Mendonça, G. Campos, P. Dias, and J. Santos. "Learning auditory space: generalization and long-term effects." *PloS one* 8.10 (2013): e77900.

[9] B. Gardner, and K. Martin. "HRFT Measurements of a KEMAR Dummy-head Microphone." MIT Media Lab Perceptual Computing – Technical Report #280, May, 1994.

978-1-5090-2096-6/16/$31.00 ©2016 IEEE