ON THE ADAPTATION TO NON-INDIVIDUALISED HRTF AURALISATIONS: A LONGITUDINAL STUDY

CATARINA MENDONÇA[1], JORGE A. SANTOS[1,2,3], GULHERME CAMPOS[4], PAULO DIAS[4], AND JOSÉ VIEIRA[4]

[1] *Departamento de Psicologia Básica, Escola de Psicologia, University of Minho, Portugal*

mendonca.catarina@gmail.com

[2] *Centro Algoritmi, University of Minho, Portugal* [3]*Centro de Computação Gráfica, Guimarães*

jorge.a.santos@psi.uminho.pt

[4] *Departemento de Electrónica, Telecomunicações e Informática, University of Aveiro, Portugal*

guilherme.campos@ua.pt, paulo.dias@ua.pt, jnvieira@ua.pt

Auralisations with HRTFs are an innovative tool for the reproduction of acoustic space. Their broad applicability depends on the use of non-individualised models, but little is known on how humans adapt to these sounds. Previous findings have shown that simple exposure to non-individualised virtual sounds did not provide a quick adaptation, but that training and feedback would boost this process. Here, we were interested in analyzing the long-term effect of such training-based adaptation. We trained listeners in azimuth and elevation discrimination in two separate experiments and retested them immediately, one hour, one day, one week and one month after. Results revealed that, with active learning and feedback, all participants lowered their localization errors. This benefit was still found one month after training. Interestingly, participants who had trained previously with elevations were better in azimuth localization and vice-versa. Our findings suggest that humans adapt easily to new anatomically shaped spectral cues and they are able to transfer that adaptation to non-trained sounds.

INTRODUCTION

Over the last decades we have witnessed a rapid growth in audio technology for the reproduction of acoustic space. The use of binaural systems that recreate the effect of the listeners' anatomical sound shading is among the most sophisticated and effective solutions. These binaural sounds are produced with specific filters, known as Head-Related Transfer Functions (HRTF), which describe the effect of head, torso and outer ear on the audio signals.

In the implementation of these auralisation techniques, the use of individualised HRTF might be preferable [1], as these filters vary considerably among different people. There are, however, limitations to the broad implementation of individualised binaural systems, mainly because HRTF measurements are time consuming and costful. With this in mind, several generic HRTF datasets were created and active efforts to find the best HRTF sets were developed (e.g. [2][3][4]). But the use of these non-individualised auralisations brought some unanswered concerns. It is assumed that listeners adapt spontaneously to new HRTFs and some data supports this assumption [5]. Wenzel and colleagues [1] compared the localization accuracy when listening to external free-field acoustic sources and to virtual sounds filtered by non-individualised HRTFs. They found an overall similarity between the results obtained in the two test situations. A similar result was found in the auralisation of speech signals [6], as most listeners obtained useful azimuth information from speech filtered with non-individualised HRTFs.

But other findings highlight that, despite some effectiveness, non-individualised sounds do not provide the same auditory experience as individualised auralisations. There is a significant increase in the feeling of presence when virtual sounds are processed with individualised binaural filters instead of generic HRTFs [7]. There are also differences in the intensity of the auditory virtual experience [8]. Comparing sound localization with arrays of speakers (twenty four or eight speaker sets) and non-individualised auralised sounds, there are significantly worse performances with the latter [9]. These contradictory results, where some findings reveal good spaciousness of the non-individualised auralisations, whereas others reveal clear differences, might be explained in light of the quality of the HRTF datasets. However, an alternative explanation could be found in the different adaptation processes that each listener undergoes to the new spectral cues of the auditory space.

From a neurological perspective, some recent data have demonstrated that humans can learn to localize with altered spectral stimulation [10]. It was found that our ability to localize sounds is experience-based, as the brain associates specific auditory cues with locations in the world. Some recent experiments [11][12] have shown that by physically altering a listener's pinnae the elevation localization ability is impaired, but in less than a month this ability is restored. This finding is regarded as evidence of the brain plasticity and ability to change with altered stimulation [13].

In the study here reported we were interested in the adaptation process to non-individualised HRTF-based auralisations and in the long-term effects of this adaptation.

Some authors have already proposed that the perception of spatial sounds with non-individualised HRTFs might be affected by perceptual learning processes, which would explain the great variability among different subjects [8] and the decrease of errors as subjects adapt [14]. In a previous paper [15] we have demonstrated that listeners do learn to localize non-individualised auralisations. There is no accuracy improvement without feedback in short periods of time, but with controlled training subjects significantly improve their performances. Also, common front-back errors frequently found in azimuth localization are reduced after a short training [16]. Crucially, this improvement is not limited to the trained sounds but to other untrained stimuli, suggesting that listeners might learn the new HRTF as a whole, not just specific new spectra or interaural differences for each sound location.

Here, we intended to explore the long-term effects of this learning process. We were interested in specifically assessing if the localization improvement was limited to the period immediately after training or if this effect lasted over time. For that purpose, we designed a longitudinal study where subjects learned elevation and azimuth localization and were afterwards consecutively tested up to one month after training.

2 METHOD

The study reported here comprised two experiments, one with stimuli varying in azimuth and the other in elevation. Each of those experiments was then composed of six smaller experiments at different points in time. As all the experimental methodology was similar, all experiments are described together in this section.

2.1. Participants

Four naïve and inexperienced subjects participated in the experiment. They all had normal hearing, verified by standard audiometric screening at 500, 750, 1000, 1500 and 2000 Hz. All auditory thresholds were below 15 dB HL and none had interaural sensitivity differences above 5 dB.

2.2. Stimuli

The stimuli consisted of auralised white noise sounds. In the azimuth experiments there were ten stimuli ranging from front to right at 10º intervals: 0º (front), 10º, 20º, 30º, 40º, 50º, 60º, 70º, 80º, and 90º (right). All these stimuli had fixed elevation (0º) and distance (1m). In the elevation experiments, the same spatial intervals were used, but with fixed azimuth (0º) and variable elevation, ranging from front to top of the head: 0º (front), 10º, 20º, 30º, 40º, 50º, 60º, 70º, 80º, and 90º (top).

In the auralisation, the original white noise sounds were convolved with the HRTF pair corresponding to the desired source position. The HRFT database used was the CIPIC [3].

Sounds were reproduced with a Realtec Intel 8280 IBA sound card, and presented through a set of Etymotics ER-4B MicroPro flat-response in-ear earphones.

2.3. Procedure

Both the azimuth and the elevation experiments started with a pre-test. In the pre-test, all sounds were presented pseudo-randomly with ten repetitions each. Participants had to indicate, on a continuum displayed in a touch screen (Figure 1), the point in space where they estimated the sound source to be. In the azimuth experiment, responding in the top area or the semicircle would mean "front" and in the right area "right". In the elevation experiment the top area corresponded to "top" and the right area to "front".

Each trial had the duration of 3 sec, with 2 sec interval between stimuli.

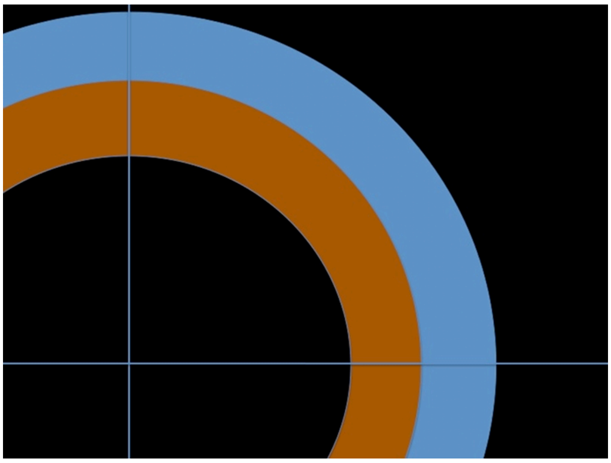

Figure 1. Pre-test and Post-test touch screen interface

After the pre-test, participants engaged in a training period. In the azimuth experiment the trained sounds were 0º, 30º, 60º and 90º (see white areas in figure 3). In the elevation experiment there were the same training areas (0º, 30º, 60º and 90º) but ranging in elevation.

The training followed the same steps as applied in our previous work [15]:

1. Active Learning: Participants were presented with sound player interface where they could hear the training sounds at their will by pressing in the corresponding area (figure 2). They were informed that they had five minutes to learn the position of each sound and that afterwards they would be tested.

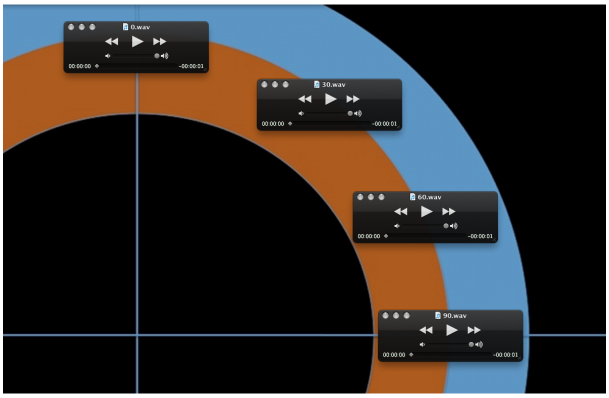

Figure 2. Active training interface.

2. Passive Feedback: After the active learning, participants heard the training sounds and had to point their location on a touch screen. After each trial, they were shown the correct answer. The passive feedback period continued until participants could answer correctly in 80 percent of the trials in the azimuth experiment, or 70 percent of the trials in the elevation experiment.

Here, each sound had the duration of 3 sec and the interstimulus interval was 4 sec.

When training period ended, participants performed the post-tests experiments, equal to the pre-test. There was a post-test immediately after training, and then another one hour, one day, one week and one month later.

Half participants trained elevation first and half participants trained azimuth first. All experiments took place in a quiet room in total darkness.

3 RESULTS

3.1. Azimuths

The individual results of the azimuth experiments are presented in figure 3. Results were analysed in terms of the localisation errors, the distance between the listeners' responses and stimuli positions, expressed in degree units.

The baseline error, in the pre-test session, was in average 17.75º. After training, participants dropped, in average, 4.4º in response error. All participants gained accuracy after training. Listener 1 and listener 2 had participated in the elevation training prior to the azimuth experiment, whereas listener 3 and 4 were totally inexperienced. Interestingly, there was a clear difference in performance between both pairs of subjects. Listeners 1 and 2 started with the lowest baseline errors (L1=14.8º, L2=17.5º) and listeners 3 and 4 with the highest (L3=18.9º, L4=19.8º). Also, the experienced subjects dropped more in localization error after training (L1=4.6º, L2=6.6º) than inexperienced subjects (L3=4.3º, L4=2.1º). These results might be evidence of a global learning of the new ear model that transfers across tasks without direct training.

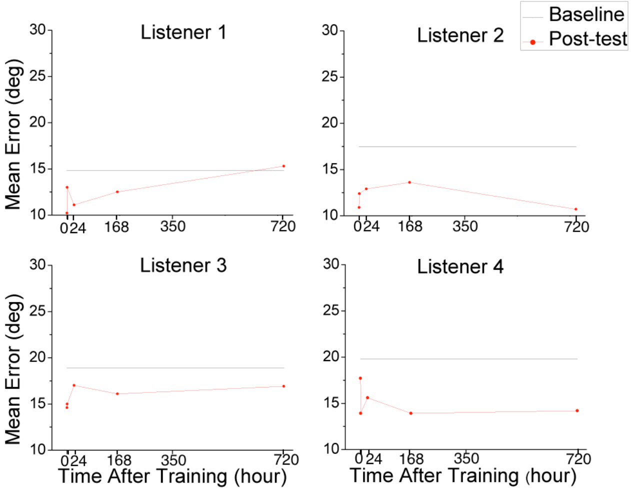

Figure 3. After training azimuth discrimination error.

Averaging the results from all participants, it is observed that the training benefit was still found in all post-tests. One month after the initial training (720th hour), the mean decrease in localisation error was 3.48º. Only one subject, listener 1, did not reveal this benefit at the last post-test. It is noteworthy that this listener had achieved a remarkable accuracy and that, at this last test, his performance was still near the performances of L2 and L3 (15.3º).

Overall, there was a clear training benefit with persistent effects over time.

3.2. Elevations

Elevation baseline localisation accuracy was, as expected, worse than that of azimuths. The average baseline was 29.47º of error, very close to the level of response at random (33º).

After training, participants reduced, in average, 7.13º in response error, more than in the azimuths experiment (4.4º). All participants showed a training benefit (figure 4). Here, listeners 3 and 4 had previous training by having participated in the azimuths experiment. They were also the participants with the lower baseline errors (L3=25.2, L4=27.1º). Listeners 1 and 2, being totally inexperienced, had baseline performances at chance (L1=34.8, L2=30.8).

All participants dropped similar levels of error (L1=8.9º, L2=7.6º, L4=7.7º), with the exception of listener 4 (L4=4.3º), who had already started with lower localisation levels.

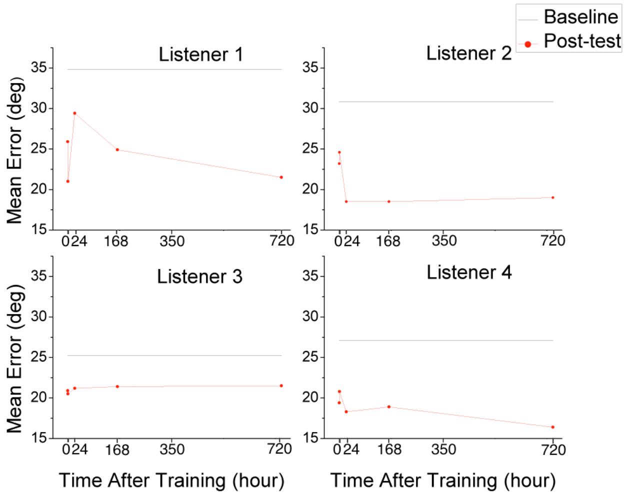

Figure 4: After training elevation discrimination error.

All participants revealed an enduring training effect. The average error reduction one hour after training was 7.75º; one day later (24 hour) it was 7.6; one week later (168 hour) it was 8.55º; and one month later (720 hour) there was still a 9.88 error reduction.

In sum, there was a clear training benefit with persistent effects over time for all subjects. Again, the previous unrelated training revealed to be beneficial to the localisation accuracy levels.

4 CONCLUSION

In the experiments here presented we intended to analyse the effects of training non-individualised HRTF-based sounds over time. In two separate groups of experiments, we trained listeners in azimuth and in elevation localisation. Subjects were afterwards retested immediately, one hour later, one day, one week and one month later.

Subjects varied greatly in their accuracy levels, but overall results revealed that both in azimuth and in elevation discrimination there was a lasting improvement effect over the localisation with the new HRTF cues. Furthermore, in both experiments, inexperienced subjects performed worse than those who had participated in a previous unrelated training (having trained elevations before azimuth testing or vice-versa). This result is compatible with our previous findings [15] where the benefit of training was not limited to the trained stimuli but also to sounds at other points in space. It might be argued that the cues are learned in direct association with the stimulus, but they are encoded as a wider frame, which is then applied to other new stimuli. The fact that this new HRTF is remembered for longer periods of time is consistent with the data from Van Wanrooij and Van Opstal [11]. It is therefore arguable that listeners can learn and use simultaneous HRTF algorithms, much like learning simultaneous languages.

This effect brings good news to the audio industry, as the initially poor and unstable results obtained by new HRTF listeners seem to be easily overcome and with lasting results, given the appropriate training.

_ ACKLOWLEDGEMENTS

This work was supported by the Portuguese Foundation for Science and Technology with the research grant SFRH/BD/36345/2007 and the funded project PTDC/EEA-ELC/112137/2009.

_ REFERENCES

[1] Wenzel, E. M., Arruda, M., Kistler, D. J., Wightman, F. L. (1994). Localization using nonindividualized Head-Related Transfer Functions *Journal of the Acoustical Society of America*, 94, pp. 111-123.

[2] B. Gardner, K. Martin. *HRTF Measurements of a KEMAR Dummy-Head Microphone*. url: http://sound.media.mit.edu/resources/KEMAR.html visited - June 2010.

[3] Algazi, V. R., Duda, R. O., Thompson, D. M. (2001). The CIPIC HRTF database. *Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics, New York, October 21-24, pp*. 99-102.

[4] Yairi, S., Iwaya, Y., Suzuki, Y. (2008). Individualization feature of head-related transfer functions based on subjective evaluation. *Proceedings of the 14th International Conference on Auditory Display, Paris, June 24-27*.

[5] Loonis, J. M., Klatzky, R. L., Golledge, R. G. (1999). Auditory distance perception in real, virtual and mixed environments, In *Mixed Reality: Merging Real and Virtual Worlds*, Y. Ohta, H. Tamura, Eds. Tokio: Ohmsha.

[6] Hiekkanen, T., Mäkivirta, A., Karjalainen, M. (2009). Virtualized listening tests for loudspeakers. *Journal of the Audio Engineering Society, 57(4),* 237-251.

[7] Valjamae, A., Larson, P., Vastfjall, D., Kleiner, M. (2004) Auditory pressure, individualized Head-Related Transfer Function, and illusory ego-motion in virtual environments. *Proceedings of the Seventh Annual Workshop in Presence*, Spain.

[8] Begault, D. R., Wenzel, E. M. (1993). Headphone localization of speech. *Human Factors*, 35(2), pp. 361-376, 1993.

[9] Ballas, J. A., Brock, D., Stroup, J. F., Fouad, H. (2001). The effect of auditory rendering on perceived movement: Loudspeakers density and HRTF. *Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29-August 1.*

[10] King, A. J. (2002). Neural plasticity: How the eye tells the brain about sound location. *Current Biology, 12,* R393-395.

[11] Van Wanrooij, M. M., Van Opstal, A. J. (2005). Relearning sound localization with a new ear. *The Journal of Neuroscience, 25(22),* 5413-5424.

[12] Van Wanrooij, M. M., Van Opstal, A. J. (2004). Contribution of head shadow and pinna cues to chronic monoaural sound localization. *Journal of Neuroscience, 24,* 4163-4171.

[13] Knudsen, E. I. (2002). Instrumented learning in the auditory localization pathway of the barn owl. *Nature, 417,* 322-328.

[14] Asano, F., Suzuki, Y., Stone, T. (1990). Role of Spectral cues in median plane localization. *Journal of the Acoustical Society of America*, 80, 159-168.

[15] Mendonça, C., Santos, J. A., Campos, G., Dias, P., Vieira, J., & Ferreira, J. (2010). On the improvement of auditory accuracy with non-individualized HRTF-based sounds. *Journal of the Audio Engineering Society, Proceedings of the 129th AES Convention,* San Francisco.

[16] Zahorik, P., Bangayan, P., Sundareswaran, V., Wang, K., & Tam, C. (2006). Perceptual recalibration in human sound localization: Learning to remediate front-back reversals. *Journal of the Acousctical Society of America, 120,* 343-359.