On the improvement of localization accuracy with non-individualized HRTF-based sounds

Catarina Mendonça[1], *AES Member*, **Guilherme Campos**[2], *AES Member*, **Paulo Dias**[2], **José Vieira**[2],
AES Fellow, **João P. Ferreira**[1], **Jorge A. Santos**[1]
Catarina.Mendonca@ccg.pt

University of Minho: School of Psychology; Centro Algoritmi; Centro de Computação Gráfica, Guimarães, Portugal

University of Aveiro, Department of Electronics Telecomunications and Informatics, Aveiro, Portugal.

Auralization is a powerful tool to increase the realism and sense of immersion in Virtual Reality environments. The Head Related Transfer Function (HRTF) filters commonly used for auralization are non-individualized, as obtaining individualized HRTFs poses very serious practical difficulties. It is therefore extremely important to understand to what extent this hinders sound perception. In this paper, we address this issue from a learning perspective. In a set of experiments, we observed that mere exposure to virtual sounds processed with generic HRTF did not improve the subjects' performance in sound source localization, but short training periods involving active learning and feedback led to significantly better results. We propose that using auralization with non-individualized HRTF should always be preceded by a learning period.

0 INTRODUCTION

Binaural auralization consists in the process of spatializing sounds. The aim is to accurately simulate acoustic environments and provide vivid and compelling auditory experiences. It has applications in many fields; examples range from flight control systems to tools for helping the visually impaired. It also has a strong potential for virtual reality (VR) applications and in the entertainment industry. This acoustic simulation should take into account the influence of the listener's anatomy over the sounds. In fact, the interaction of sound waves with the listener's body – particularly torso, head, pinnae (outer ears) and ear canals – has extremely important effects in sound localization, notably interaural time and level differences (ITD and ILD, respectively), the main cues for static source localization. Such effects can be measured as a binaural impulse response for the corresponding source position, known as *Head Related Impulse Response (HRIR)*, or by its Fourier transform, the *Head Related Transfer Function* (HRTF). It is possible to appropriately spatialize headphone-delivered sounds by processing anechoic recordings of the source material through the HRTF filters corresponding to the desired virtual source position [1][2][3][4].

Since they depend on anatomic features such as the size and shape of head and ears, HRTFs vary considerably from person to person. From this fact emerged two distinct auralization approaches: individualized HRTFs, made for each listener from their own individual features, and generic/averaged HRTFs. Given the between-subject variability in HRTFs, it is arguable that all spatial audio simulations should use individualized HRTFs. However, this is extremely difficult to obtain in practice; HRTF recordings are effortful and expensive, requiring anechoic rooms, arrays of speakers (or accurate speaker positioning systems), miniature microphones, and specialized software and technicians. Alternatively, generic sets are measured on manikins or head-and-torso systems equipped with artificial pinnae designed to approximate as best as possible an 'average' human subject. Several additional efforts exist to minimize the differences between generic and individualized HRTFs [1,5,6], and some mixed models have been proposed [7]. However, the debate between the merits and trade-offs of indivisualised/generic auralisations still persists.

On the one hand, there is still not enough data on the efficiency of the generic (non-individualized) HRTFs in replacing the individualized ones. Not all information in an indivisualized HRTF is perceptually relevant [8]. It has been suggested that satisfactory auralization can be obtained using generic HRTFs [6]. Wenzel et al. [9] compared the localization accuracy when listening to external free-field acoustic sources and to virtual sounds filtered by non-individualized HRTFs. Several front-back and up-down confusions were found, but there was overall similarity between the results obtained in the two test situations. A similar result was found in the auralization of speech signals [10], as most listeners were able to obtain useful azimuth information from speech filtered with non-individualized HRTFs.

On the other hand, there are indications that the listening effects of individualized HRTF-based systems do differ from the generic ones [11,12]. There is a significant increase in the feeling of presence when virtual sounds are processed with individualized binaural filters instead of generic HRTFs [13]. In a study that compared real life listening with real head recordings and artificial head recordings [14], it was found that localization accuracy with recordings is worse than in real life, and that artificial heads are worse than real head recordings. Interestingly, there was a clear learning effect over the period of five days. There had been some previous suggestions that the perception of spatial sound with non-individualized HRTFs might change over time. Begault and Wenzel [10] observed several individual differences, which suggested that some listeners were able to adapt more easily to the spectral cues of the non-individualized HRTFs than others. Asano et al. [15] claimed that reversal errors decreased as subjects adapted to the unfamiliar cues in static anechoic stimuli. Jie and collaborators [16] argued that there were listening improvements with time in loudspeaker displayed sounds convolved with non-individualized HRTFs.

In this context, our primary research question in this paper is: can humans learn to accurately localize sound sources processed with HRTF sets different from their own? There is evidence that the mature brain is not immutable, but instead holds the capacity for reorganization as a consequence of sensory pattern changes or behavioral training [17]. Shinn-Cunningham and Durlach [18] trained listeners with “supernormal” cues, which resulted from the spectral intensification of the peak frequencies. With repeated testing, during a single session, subjects adapted to the altered relationship between auditory cues and spatial position. Hofman [19] addressed the consequences of manipulating spectral cues over large periods of time (19 days), adapting molds to the outer ears of the subjects. Elevation cues (which, in static listening, depend exclusively on monaural cues) were initially disrupted. These elevation errors were greatly reduced after several weeks, suggesting that subjects learned to associate the new patterns with positions in space.

The broad intention of this study was to assess how training may influence the use of non-individualized static HRTFs. Our main concern was assuring that users of such generically spatialized sounds become able to fully enjoy their listening experiences in as little time as possible.

Three experiments were designed to answer the questions: *Do listeners spontaneously improve accuracy without feedback in short periods of time?* (experiment 1); and *Can the adaptation process be accelerated by applying feedback?* (experiments 2 and 3).

1 GENERAL METHODS

1.1 Participants

The main experiment comprised a series of successive In all experiments, only naïve and inexperienced young adults were used. They all had normal hearing, verified by standard audiometric screening at 500, 750, 1000, 1500 and 2000 Hz. All auditory thresholds were below 10 dB HL and none had significant interaural differences (threshold differences were below 5 dB HL at target frequencies). There were always 4 participants for each experiment. In the last two experiments participants were the same, half of the participants started with Experiment 2 and half started with Experiment 3.

1.2 Stimuli

The stimuli in all experiments consisted of pink noise sounds auralized at several positions in space. The original (anechoic) sound was convolved with the HRTF pair corresponding to the desired source position. The resulting pair of signals – for the left and the right ear – was then reproduced through earphones. No fade-in/fade-out were used in the start/end of the signal.

The HRTF set were recorded using a KEMAR dummy head microphone at the Massachusetts Institute of Technology [20]. Sounds were reproduced with a Realtec Intel 8280 IBA sound card, and presented through a set of Etymotics ER-4B MicroPro in-ear earphones.

All sounds were presented pseudo-randomly: they were randomized assuring the same event number for all stimuli. Each stimulus had a three second duration, with one second of inter-stimulus interval.

2 EXPERIMENT 1

This experiment intended to assess the localization accuracy of inexperienced subjects as they became gradually more familiarized with the non-individualized HRTF processed sounds. We tested their ability to locate different azimuth sounds in the horizontal plane in 10 consecutive experimental sessions (blocks), while not providing any feedback on the accuracy of their responses. We analyzed the evolution of the subjects’ performance across blocks.

2.1 Method

Sounds were auralized at 8 different azimuths: 0º (front), 180º (back), 90º (left and right), (45º left and right), and 135º (left and right). They had constant elevation (0º) and distance (1m).

There were 10 blocks, each one with 10 stimuli repetitions. Therefore, each block presented a total of 80 sounds per subject. Participants were told to indicate the perceived sound source location for each stimulus.

The answers were recorded by selecting, on a touch screen, one of the eight possible stimulus positions.

2.2 Results

The average accuracy of azimuth localization was above chance (65% correct answers) in all cases, but no ceiling performances were observed. The left and right 90º sounds were on average the most accurately located, with a correct response rate of 78% (see figure 1). Similarly to what had been found in previous studies [8][9], there were several front-back confusions that account for the lower accuracy at 0º (62% correct answers), 180º (43%), left/right 45º (60%) and left/right 135º (69%).

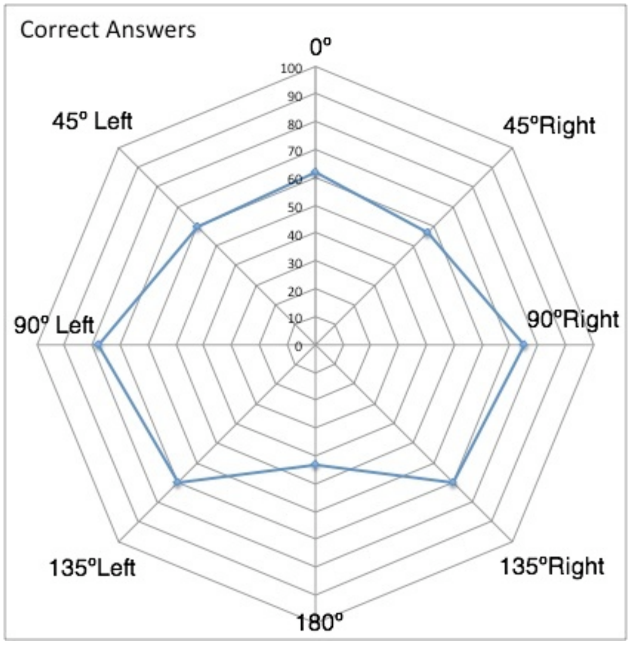

Figure 1. Percentage of correct answers by azimuth (400 trials per value).

Despite the data presented in figure 1, we should not consider the analysis by azimuth representative, as each listener revealed different tendencies and biases. Indeed, there were lateral asymmetries in accuracy, as well as no fixed rule in the error typology. For instance, some participants failed to answer correctly to the 45º stimuli due to front-back confusions, whereas others failed due to confusions with the 90º sounds. To further analyze these effects a more comprehensive study, with more participants, would be needed.

Analyzing the average participants' performance along time (figure 2), we see that the overall accuracy remained constant. There were individual differences between participants. Listener 1 was the least accurate (50.4% correct answers), listeners 2 and 3 performed near average (61.9% and 71.1%, respectively) and listener 4 had the best azimuth localization performance (85.1%).

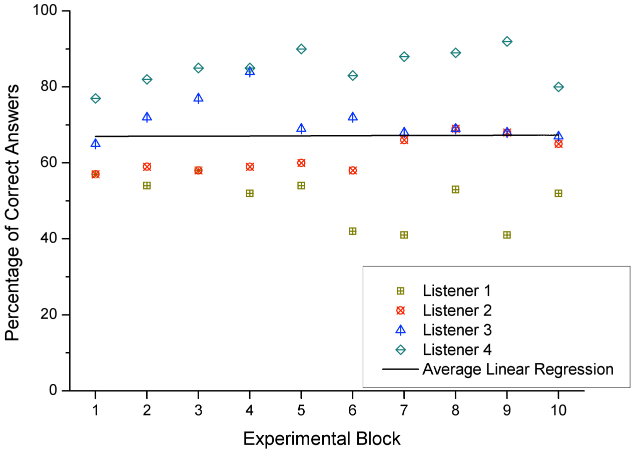

Figure 2. Percentage of correct answers by experimental block and linear regression (80 trials per dot).

The linear regression results revealed a slope coefficient close to zero (0.04), meaning almost no change in the percentage of correct responses. The concatenated correlation values revealed that indeed the experimental block number did not account for the listeners' accuracies (r^2=0.00, p=0.66). The analysis of each individual participant's correlations revealed that none obtained a significant effect of block number over correct responses. This result was further supported by hypothesis testing ($F\text{ANOVA}_{9,3}$=0.087, p=0.99), which revealed no significant interactions between each block and each listener's correct answers.

2.3 Discussion

Our results reveal that naïve participants are able to localize sounds at several azimuths. However, this ability is neither high nor consistent among subjects. Furthermore, throughout the exposure blocks, their accuracy does not evolve, leading to the conclusion that simple exposure is not enough for significant localization improvement in short periods of time.

In view of these conclusions, a second experiment was developed where, in the same amount of time, listeners were trained to identify sound source locations.

3 EXPERIMENT 2

In experiment 2, we tested the participants' accuracy in localizing sounds at several azimuths before and after a short training program. In this program, we selected only a small number of sounds and trained them through active learning and response feedback.

3.1 Method

All stimuli were auralized varying in azimuth, with elevation (0º) and distance (1m) fixed. Azimuths ranged from the front of the subjects head to their right ear, spaced at 6º intervals (from 6º left to 96 º right). Only these azimuths were used, aiming to assure that other effects such as front-back biases (like a subject's tendency to perceive sounds in the back area rather than in the front) and individual lateral accuracy asymmetries (like a subject's tendency to be more accurate for the right sided sounds) did not emerge, as they were not the focus of our study. Stimuli did not range solely from 0º to 90º, but from 6º to 96º, to avoid reducing response options, which would artificially increase the accuracy at these azimuths.

3.2 Procedure

The experiment started with a pre-test. In the pre-test, all sounds were presented with 4 repetitions each. Participants had to indicate, on a continuum displayed on a touch screen (figure 3A, blue area), the point in space where they estimated the sound source to be.

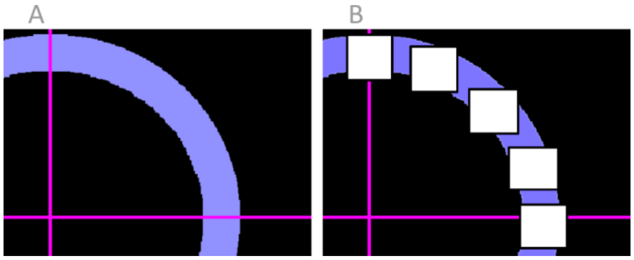

Figure 3. Touch screen in the pre-test and post-test (A). Touch screen in the training program (B).

After the pre-test, participants engaged in a training period. The trained sounds corresponded to the frontal (0º), lateral (90º) and three intermediate azimuths (21º, 45º and 66º) (see white areas in figure 3B).

The training conformed to the following steps:

Active Learning: Participants were presented with a sound player where they could hear the training sounds at their will. To select the sounds, there were several buttons on the screen, arranged according to the corresponding source position in space. The participants were informed that they had 5 minutes to practice and that afterwards they would be tested.

Passive Feedback: After the 5 minutes of active learning, participants heard the training sounds and had to point their location on a touch screen (figure 2B). After each trial, they were told the correct answer. The passive feedback period continued until participants could answer correctly in 80 percent of the trials (5 consecutive repetitions of all stimuli with at least 20 correct answers).

When the training period ended, participants performed a post-test, an experiment equal to the pre-test for comparison purposes.

3.3 Results

3.3.1 Pre-Test

Results from the pre-test and post-test sessions are displayed in figures 4 and 5.

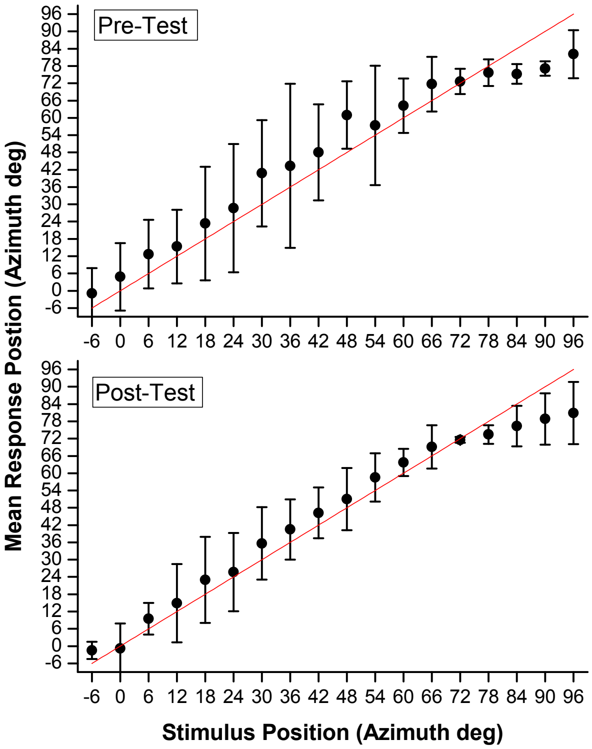

Figure 4. Mean azimuth localization tendencies for each stimulus and standard deviations between subjects (120 trials per value).

Observing the average subjective localization for each stimulus (figure 4), we find there were no major shifts in the perceived location of each sound. In the pre-test there were greater azimuth mean value deviations in the central area, where sound tended to be perceived more to the right; and in the right area, where the opposite effect occurred. Standard deviations were much larger in the central area, which might reflect greater uncertainty and not a significant shift in the perceived stimulus location.

In figure 5, orange and purple bars display the average distance (in degrees) to the given stimulus position. Gray bars display the mean theoretical error (in degrees) that would be obtained if participants responded randomly.

Analyzing the pre-test error results (figure 5, orange bars), we observe that azimuth localization is easier for frontal stimuli: the average error is below 5 degrees. The absence of rear stimuli which prevented any front-back confusions might help explain these results. As in experiment 1, listeners were fairly precise in identifying lateral source positions. Sounds were most difficult to locate at intermediate azimuths (between 40º and 60º). For these positions, pre-test localization was at chance level, revealing an overall inability of the subjects to accurately identify such sound positions.

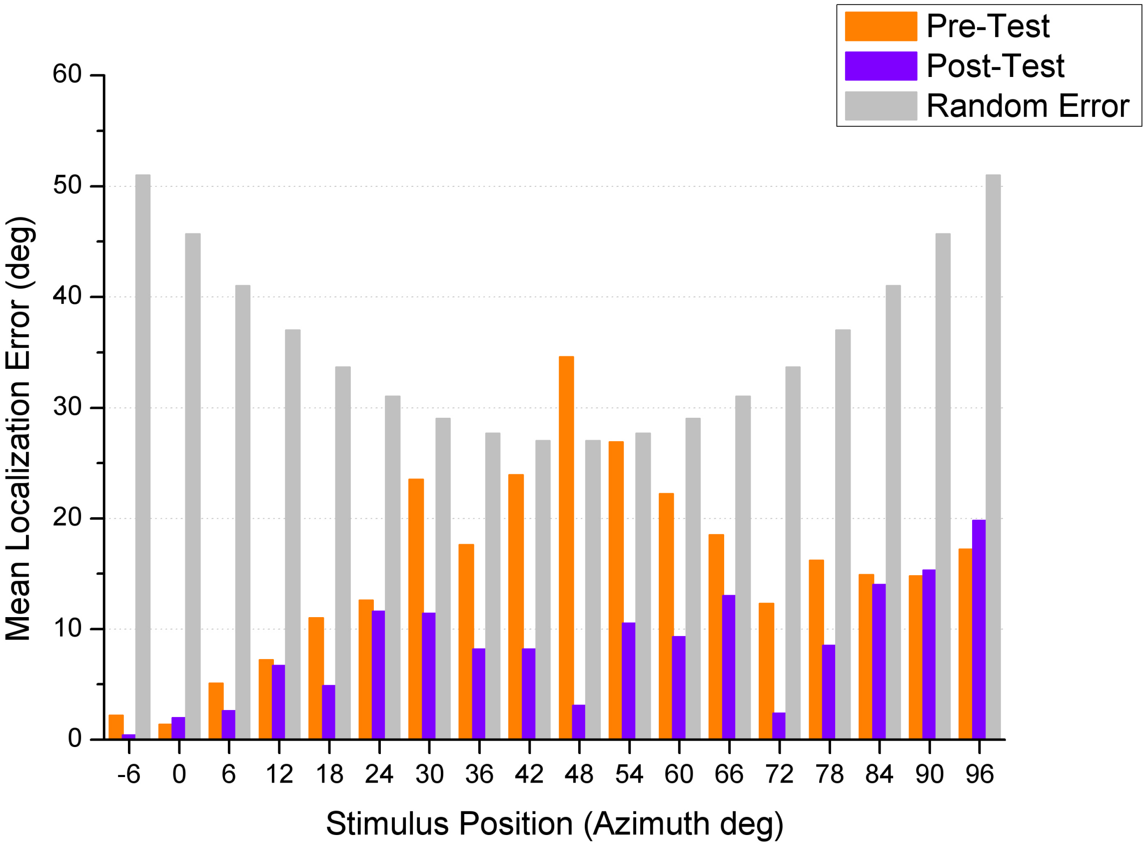

Figure 5. Average response error in the Pre-Test and Post-Test sessions (120 trials per value), and theoretical error level if listeners responded randomly.

On average, participants missed the stimulus position in the pre-test by 15.67º.

It is noteworthy that it was precisely in the central area that larger standard deviations occurred, which is consistent with the larger errors already found for this area in figure 4.

3.3.2 Training Period

The training sessions were very successful for all participants. All took less than 30 minutes and, in average, they lasted 22 minutes.

Learning curves are displayed in figure 6, where individual azimuth localization accuracy is plotted as a function of the time elapsed since the start of the training period.

All participants reached the 80% criterion. Despite the differences in learning velocity, a monotonic progression was observed for all of them.

3.3.3 Post-Test

The post-test results (figure 5, purple bars) revealed a large error reduction of 7.23º on average, from 15.67º in the pre-test to 8.44º in the post-test. Despite individual differences, all participants revealed similar learning effects.

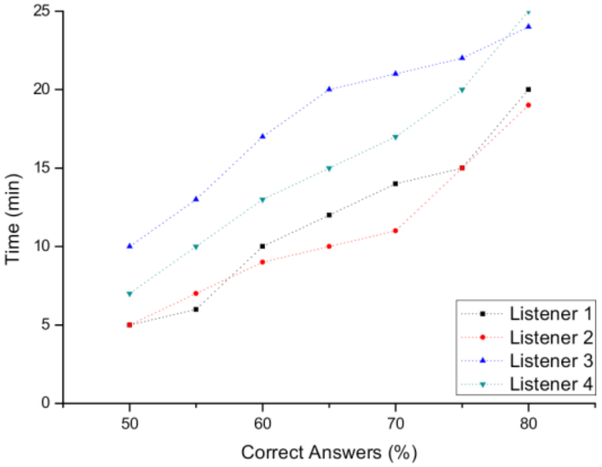

Figure 6. Individual accuracy evolutions in the azimuth localization training sessions.

This difference was statistically significant in a paired samples T-test ($t_{(287)}=14.94$, $p \leq 0.001$). The error reduction was most expressive in the intermediate azimuths, where the average error decreased 20 degrees.

Analyzing the trained azimuths (0º, 21º, 45º, 66º, 90º), we observe that performance enhancement was substantial not only for these stimuli, but also for others, not trained. As an example, the best error reduction was obtained with the 48º azimuth, an untrained stimulus. In contrast, the 90º azimuth, a trained one, revealed similar results in both sessions.

Looking at average localization tendencies (figure 5) in the post-test listeners became slightly more precise, especially for the right area azimuths and variability across subjects was reduced in the central area.

3.4 Discussion

In this experiment we trained subjects in azimuth localization. We found that listeners learned fast, in a small amount of time. They improved their localization ability not only for the trained azimuths, but also for others. These findings allow us to conclude that the trained localization abilities for some stimulus positions are generalized to other, untrained, auditory positions.

4 EXPERIMENT 3

In experiment 3, an elevation localization task was carried out using the same methodology as in experiment 2. Static elevation is known to be perceived less accurately than azimuth, probably because it does not benefit from as many binaural cues as azimuth. This experiment was designed to investigate whether or not the learning effect found in experiment 2 could be attributed to an improved interpretation of the binaural information contained in the HRTFs.

4.1 Method

In this experiment, the stimuli varied in elevation, but not in azimuth (0º) or distance (1m). They ranged from the front of the listeners' head (0º in elevation) to the top (90º in elevation) in 10º intervals. Participants were aware that no back stimuli were present, but no instruction was provided regarding stimuli below 0º.

4.2 Procedure

Experiment 3 followed the same procedure as experiment 2.

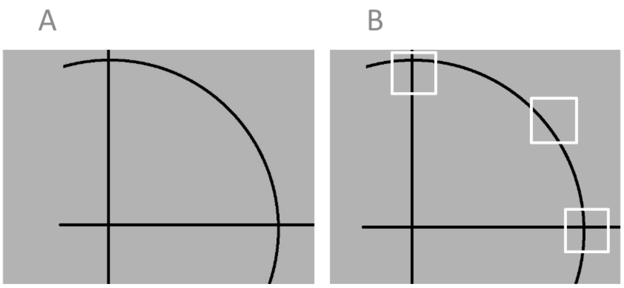

Figure 7. Touch screen in the pre-test and post-test (A). Touch screen in the training period (B).

In the training period, the sounds were positioned at elevations of 0º, 40º and 90º. Figure 7 shows the touch screen used in the pre-test and post-test sessions (A), as well as the touch screen with the 3 defined elevations, which were trained (B).

4.3 Results

4.3.1 Pre-Test

Figure 8 presents the average distance (in degrees) between the subjects' answers and the stimuli elevations in the pre and post-test sessions. It also shows the theoretical errors that would be obtained if subjects responded at chance. These bars account for mean error if subjects responded only between 0º and 90º elevation. In fact, participants did sometimes answer below 0º, making response distribution asymmetric. Nevertheless, only symmetric predictions are presented.

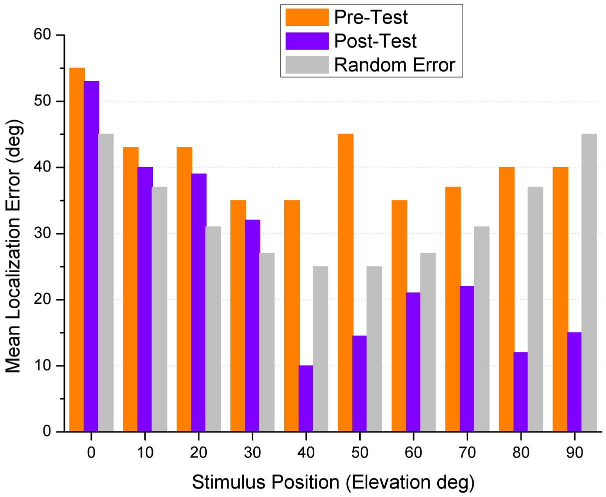

Figure 8. Average response error in the Pre-Test and Post-Test sessions (120 trials per value), and theoretical response errors if listeners responded randomly.

In the pre-test session, the average error was of 40.8º, close to random error. The subjects were unable to identify the target sound position at any elevation; the worst results were in the frontal (0º) stimuli (55º average error). Overall, participants were less accurate in estimating a sound position in elevation than in azimuth.

Regarding where each elevation sound was perceived in space (figure 9), we observe that prior to training sounds were not accurately located, falling into median locations with large standard deviations. Standard deviations were larger for frontal stimuli, but all sounds tended to be perceived at higher elevations.

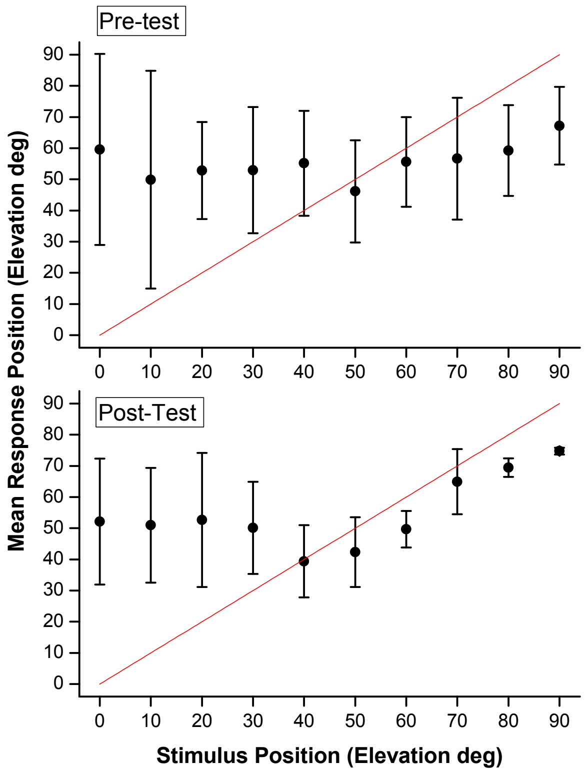

Figure 9. Mean elevation localization tendencies for each stimulus and standard deviations between subjects (120 trials per value).

4.3.2 Training Period

Training sessions were faster than those of experiment 1, as there were only 3 trained elevations. On average, they took 17 minutes (Figure 10).

Only one subject (listener 3) did not evolve as expected. After 10 minutes testing, this subject was still making excessive mistakes, and was allowed a second active learning phase (5 minutes), after which the 80 percent accuracy was rapidly achieved.

4.3.3 Post-Test

The post-test results were better than those of the pre-test for all subjects (figures 8 and 9). This difference was significant in a paired samples T-test ($t_{(159)}$=16.678, $p \leq 0.001$) The average error decreased 14.75 degrees, to a mean of 26.5º (figure 8), an effect larger than found in experiment 2. The training effect was most expressive for the upper stimuli, namely at 80º, 40º and 50º elevations. Among these stimuli, the only trained one was at 40º. On the other hand, errors for sounds at 0º elevation, a trained stimulus, revealed no significant decrease in the post-test session. Similarly to what was found in experiment 2, training was effective and generalized well to other stimuli.

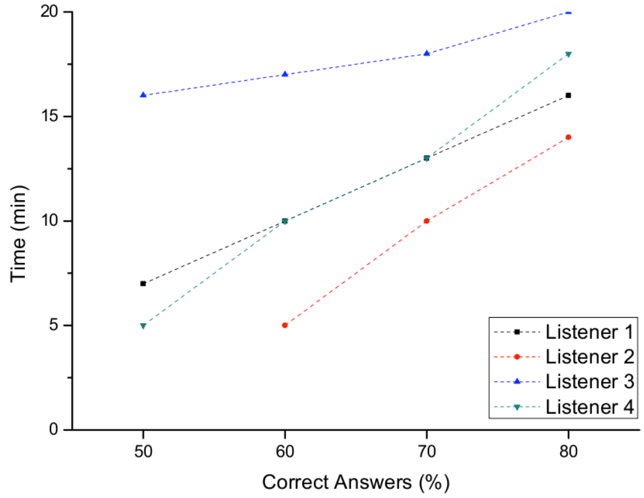

Figure 10. Individual accuracy evolutions in the elevation training sessions.

Regarding where sounds were perceived in space (figure 9), there was an improvement of localization accuracy for all stimuli along with a standard deviation decrease, but for sounds below 40º.

4.4 Discussion

In this experiment we trained subjects in an elevation localization task. As in the second experiment we found that listeners learned quite fast, in a small amount of time. There was an overall better localization accuracy for the upper elevations. Sounds at lower elevations did not improve with training. This result might be interpreted as a general inability to accurately interpret these sounds, but it might as well be a result of a methodological artifact. As the test subjects were not aware that no stimuli existed below 0º, some responded below, artificially elevating the response distribution and therefore mean error. An additional experiment controlling methodological awareness would de needed to obtain conclusive results.

In general, listeners improved their localization ability not only for the trained elevations, but also for others. These findings bring further support to the assumption that the learning achieved for specific sound positions might be transferred to other, untrained, positions.

5 FINAL DISCUSSION

In this paper, we were specifically interested in better understanding the evolution in perceptual accuracy as a subject familiarizes with non-individualized HRTFs. We intended to understand if listeners adapt spontaneously without feedback in a reasonably short time and/or if we could somehow accelerate the adaptation process.

In experiment 1, we addressed the listeners' adaptation process to static non-individualized azimuth sounds without feedback. Throughout 10 short experimental consecutive sessions, we measured the percentage of correct answers in position identification. Results revealed an overall absence of performance improvement in all subjects. We concluded that simple exposure is not enough for significant accuracy evolution to be achieved in short periods of time. Such exposure learning had been claimed in previous works [9][12][13], in an attempt to explain individual differences in accuracy results. Our results did not reveal such effects. Adaptation was, however, demonstrated before in [16], but over wide periods of time (weeks) and with spatial feedback, as participants of those experiments carried the molds inside their ears in their daily lives during the whole period.

Pursuing the intent of preparing untrained listeners to take full advantage of non-individualized HRTFs, we designed a second experiment, where subjects could train with sample sounds in a short program combining active learning and feedback. In a pre-test, participants revealed good localization abilities for frontal stimuli, but performed very poorly in the intermediate (40º to 60º) azimuths. After the training sessions, in a post-test, all azimuths were identified above chance, with results significantly better than the pre-test ones. More importantly, the training benefit was observed not only for the trained sample azimuths, but was also generalized to other stimulus positions. In an attempt to interpret such results, one might argue that an overall learning of the new HRTF-based cues took place, and was then applied to the other untrained stimuli.

In experiment 3, we tested the same training program, with stimuli varying in elevation and with fixed azimuth. Elevation alone is known to be poorly perceived in space, when compared to azimuth, mostly because it relies less on binaural cues as ITD and ILD values are fixed. Results in the pre-test of this experiment revealed poor source localization ability at almost all elevations, particularly from 0º to 50º. Indeed, with unfamiliar HRTF filters, auralized sounds carried little elevation information for the untrained subjects. A large difference was found in the post-test, where some localization ability arose. Again, the performance benefit was generalized across stimuli and was not restricted to the trained elevations. This finding further supports the assumption that indeed the new HRTF-shaped frequencies were learned.

Both in experiments 2 and 3, the training sessions had the approximate duration of 20 minutes. Longer training sessions might have led to better performance improvements. We stress, however, that in preparing listeners for auralized interfaces time should not be the criterion. In our sessions, each participant revealed a different profile and learned at a different velocity. Fixing a goal (such as 80% accuracy) will allow a way of assuring all listeners reach an acceptable adaptation.

In this paper we used the term localization as a conceptual tool to assess how listeners learn and adapt to non-individualized HRTF-based sounds. We addressed localization as an ability to point to a given sound source in space. Therefore, we cannot reach conclusions about the underlying neural processes and most precisely, we cannot appreciate how subjects actually formed the percept of each given sound in space. There might be pattern learning processes involved or other cognitive factors that we cannot account for. Nevertheless, we should stress that there was a benefit of training not only for the listened sounds but also for other ones. So, an adaptation or learning process did take place.

We conclude that for binaural auralization using generic HRTFs it is possible to significantly improve the auditory performance of an untrained listener in a short period of time. However, natural adaptation to static stimuli is unlikely to occur in a timely manner. Without any training, several source positions are poorly located. In view of this, we argue that testing virtual sounds processed through non-individualized HRTFs should always consider possible learning or adaptation effects.

Future studies in this field should test a range of different stimulus sounds, and also focus on the endurance of the learned capabilities over time, generalization limits, and the training effects on the final auditory virtual experience.

6 ACKNOWLEDGEMENTS

This work was supported by FCT - Portuguese Foundation for Science and Technology (SFRH/BD/36345/2007 and PTDC/TRA/67859/2006) and by FEDER funding (FCOMP-01-0124-FEDER-022674).

We thank Prof. Damian Murphy (U. York) for his availability and support on methodological issues. We also thank Prof. Pedro Arezes for his technical support.

7 REFERENCES

[1] J. Blauert (ed.). *Comunication Acoustics*. Germany: Springer-Verlag, 2005.

[2] F. Wightman, D. Kistler. “Measurement validation of human HRTFs for use in hearing research.” *Acta Psic. United with Acoustica,* 91, pp. 429-439, 2005.

[3] G. Plenge. “On the difference between localization and lateralization.” *J. Acoust. Soc. Am.*, 56, pp. 944-951, 1974.

[4] F. L. Wightman, D. J. Kistler, M. E. Perkins. “A new approach to the study of human sound localization.” in *Directional Hearing,* W. Yost, G. Gourevich, Eds. New York: Springer-Verlag, 1987.

[5] D. Hammershøi, H. Møller, "Sound transmission to and within the human ear canal." J. Acoust. Soc Am. 100(1), pp 408-427, 1996.

[6] J. C. Middlebrooks. “Individual differences in external-ear transfer functions reduced by scaling frequency.” *J. Acoust. Soc.*

Am. 106(3), pp. 1480-1492, 1999. [7] J. M. Loonis, R. L. Klatzky, and R. G. Golledge. “Auditory distance perception in real, virtual and mixed environments”, in *Mixed Reality: Merging Real and Virtual Worlds*, Y. Ohta, H. Tamura, Eds. Tokio: Ohmsha, 1999.

[7] K. J. Faller II, A. Barreto, M. Adjouadi. “Audmented hankel total last squares decomposition of head-related transfer functions. *J. Audio Eng. Soc.,* 58(1/2), pp. 3-21, 2010.

[8] J. Beebaart, F. Nater, A. Kohlrausch. “Spectral and spatial parameter resolution requirements for parametric filter-bank-based HRTF processing. *J. Audio Eng. Soc.,* 58(3), 2010.

[9] E. M. Wenzel, M. Arruda, D. J. Kistler, F. L. Wightman. “Localization using nonindividualized Head-Related Transfer Functions.” *J. Acoust. Soc. Am.*, 94, pp. 111-123, 1993.

[10] D. R. Begault, E. M. Wenzel. “Headphone localization of speech.” *Hum. Fact.*, 35(2), pp. 361-376, 1993.

[11] T. Papadopoulos, P. A. Nelsar. “Choice of inverse filter design parameters in virtual acoustics imaging systems.” *J. Audio Eng. Soc.,* 58(1/2), pp. 22-35, 2010.

[12] E. Blanco-Martin, F. J. Casajus-Quirós. “Objective measurement of sound event localization in horizontal and median planes.” *J. Audio Eng. Soc.,* 53(3), pp. 124-136, 2011.

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

[14] P. Minnaar, S. K. Olesen, F. Christensen, H. Møller. "Localization with binaural recordings from artificial and human heads." *J. Audio Eng. Soc.*, 49(5), pp. 323-336, 2001.

[15] F. Asano, Y. Suzuki, and T. Stone. “Role of Spectral cues in median plane localization.” *J. Acoust. Soc. Am.*, 80, pp. 159-168, 1990.

[16] J. L. Jie, Huang, S. Akira, T. Keita, T. Watanebe. “The learning effect of HRTF based 3-D sound perception with an horizontally arranged 8-loudspeaker system. *129th AES Convention, 2010, San Francisco*.

[17] C. D. Gilbert. “Adult cortical dynamics.” *Physiol. Rev.*, 78, pp. 467-485, 1998.

[18] B. G. Shinn-Cunningham, N. I. Durlach, R. M. Held. “ Adapting to supernormal auditory location cues. I. Bias and resolution.” *J. Acoust. Soc. Am.*, 103, pp. 3656-3666, 1998.

[19] P. M. Hofman, J. G. A. Van Ristwick, A. J. Van Opstal. “Relearning sound localization with new ears”. *Nat. Neurosc.*, 1, pp. 417-421, 1998.

[20] B. Gardner, K. Martin. “HRTF Measurements of a KEMAR Dummy-Head Microphone.” *MIT Media Lab Perceptual Computing – Technical Report #280, J. Audio Eng. Soc.*, 1994.

THE AUTHORS

Catarina Mendonça

Guilherme Campos

Catarina Mendonça is a Psychologist with an MSc in Cognitive Sciences, and a PhD in Experimental Psychology and Cognitive Sciences. She currently works at University of Minho as a Research Coordinator in Perception, Interaction and Usability. Her areas of expertise are human perception of multimodal displays, perception-action loops, human adaptability and user interfaces. Since 2006 she has worked in the Laboratory of Visualization and Perception, as a researcher. From 2008 to 2011 she worked as a researcher both in the School of Psychology, University of Minho, and in the Faculty of Medicine, University of Coimbra. In University of Minho, she was a Teaching Assistant from the school years 2007/2008 to 2010/2011 in the discipline of "Laboratory of Perception" and an Invited Teaching Assistant both in 2008/2009 and in 2010/2011, having lectured "Psychology of Perception". Over the last few years, she has published under the theme of human perception, and participated in several funded research projects (e.g. research project Biomotion (PTDC/SAU-BEB/68455/2006), Noiseless (PTDC/TRA/6786 59/2006), Acousticave (PTDC/EEA-ELC/112137/2009) and in international collaborations.

Guilherme Campos graduated in Electrical and Computer Engineering at the Faculty of Engineering of the University of Porto (FEUP) in 1989, with a specialisation in Control Systems. He worked for several companies, in electronic system development, factory and building automation projects and quality assurance, and became technical director of a manufacturer of electrical equipment, with voting seat at the national electrical standards committees. He gained scholarships from FCT (Portugal's Foundation for Science and Technology) to study at the University of York (UK), where he completed an MSc in Music Technology (Dept. of Music and Electronics) in 1997 and a PhD (Dept. of Electronics) in 2003 on room acoustic modelling using 3D digital waveguide meshes. He moved back to Portugal to take a post of assistant professor at the Department of Electronics, Telecommunications and Informatics (DETI) of the University of Aveiro and pursue research in audio signal processing at its affiliated Institute of Electronics and Telematics Engineering (IEETA). His main current research interests are: 3D digital waveguide modelling and its practical application to room acoustic simulation and real-time audio virtual reality; parallel computing and development of specific hardware for physical modelling; audio pattern recognition for computer-aided stethoscopy.