Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training

Alan Blum[1, 2], Brian F.G. Katz[2], Olivier Warusfel[1]
[1] *IRCAM, 1 place Igor Stravinsky 75004 Paris, France, Email: warusfel@ircam.fr, blum@ircam.fr*
[2] *LIMSI-CNRS, BP 133 – F 91403 Orsay Cedex, France, Email: brian.katz@limsi.fr*

Introduction

Spatial hearing relies on the exploitation of acoustic cues, the well known Head Related Transfer Function (HRTF), by the auditory system. Binaural synthesis is the creation of 3D virtual sound sources based on the reproduction of these acoustic cues to the listener's ears over headphones.[1,2] From these HRTFs, one can extract inter-aural time differences (ITD), principally linked with the lateralization of sound sources, as well as spectral cues, which are more concerned with elevation. A common problem is that HRTFs are extremely morphologically dependent (shape of the outer ears, head dimensions, torso) thus providing idiosyncratic cues. Therefore, methods for generating individualized HRTFs is the topic of numerous studies [3,4] in order to avoid typical binaural synthesis artifacts such as front/back confusions or angular distortions primarily in the vertical plane.[5] In fact, although we can claim high authenticity with binaural rendering, it is not always a natural situation (for example the lack of dynamic cues due to head movements in a static rendering [6]) and, even in the best conditions, some artifacts are still present. It is often found that a period of adaptation is needed to adjust to rendering anomalies, and studies using binaural spatialization mention trial sessions for the subjects to become acquainted with the technique. This is often termed the "learning effect".[5,6,7]

An interesting extension of this phenomenon is to consider the individualization problem in a reverse sense, that is, instead of adapting HRTFs to the individual, one forces the auditory system to adapt to the non-individual HRTF. There is evidence of ongoing spatial calibration in the adult human auditory system. Work by Hofman *et al* [8] showed that subjects with modified pinnae (using inserts), following a dramatical degradation of sound elevation perception, steadily reacquired localization performance over time. The point of the present study is to make further investigations on this auditory localization adaptation in the context of a virtual sound environment. Many questions can be raised, such as how to elicit adaptation, which acoustic cues are involved, what is the duration of adaptation, how long does it last, or where in the auditory system does it occur. This study focuses on adaptation to the spectral component of the HRTF, as was done by Hofman, but directed towards virtual rendering with the condition that adaptation should be rapid. Rapid auditory training for lateralization with paired audio/visual stimuli or induced plasticity with ventriloquism effect has been previously demonstrated.[9,10,11] Here we induce adaptation to non-individual HRTFs spectral cues with a quick exploration of the spatial map by an auditory-kinesthetic process. This idea is in contrast to recent studies involving spatial hearing in blind subjects which have reported that early-blind subjects, commonly believed to have better localization performance through hyper-compensation of their visual loss, exhibit less accuracy in elevation estimation, suggesting that the auditory system may require visual feedback for calibration.[12,13] Some exceptions were noted which lead to the hypothesis that auditory calibration may be achieved via other multimodal interactions. A proprioceptive feedback of auditory spatial information has been used by giving blindfolded subjects control of a virtual sound source spatialized at the hand position. The principle is that the listener associates the source position with the acoustic cues used for the binaural rendering through the constant and innate awareness of one's own hand position. This modality has the advantage of offering natural interactivity with perception/action coupling and is not limited to the visual field, thus allowing the user to explore the entire auditory sphere.

Methods

The experiment consisted of three tests: a localization test to evaluate the initial performance of the subject, an "adaptation session", followed by the localization test again. The goal is an improvement in performance between the two localization tests. Subjects were blindfolded. Binaural rendering was accomplished using IRCAM's Spat software (no room effect), with head/hand electro-magnetic tracking. In the *localization task (steps 1 & 3)* the perceived position of a spatialized sound sample was reported using a hand pointing technique which studies have shown to be more efficient [7] while also having the ecological advantage of being egocentric and natural for the user. There was no head-tracking implemented and the user indicated when the position indicated was to be recorded via a foot pedal. A 40 Hz amplitude modulated pink noise burst stimulus (sound pressure level 55 dBA measured at the ears, duration 200 ms, rise/fall time 50 ms), repeated 3 times with 1 s and then 1.5 s between bursts. 25 positions equally distributed between left/right and front/back hemispheres were randomly presented with 5 repetitions of each (125 stimuli in total, mean duration 20 min). In the *adaptation task (step 2)* a "game" was created where subjects had to search for animal sounds hidden around them. In their hand, they were given a ball which was position tracked. The listener's head position and orientation were also tracked during this phase and the sound was spatialized at the ball center. The sound consisted of alternating pink/white noise, with the alternating frequency increasing as the distance between the ball and the searched point reduced (following the principle of a Geiger counter). When the target position was found, the "detector sound" was replaced by an "animal". The subject then placed the "animal" at a reference point directly in front. In this way continual proprioceptive and auditory information were spatially aligned and the listener was forced to explore his/her whole sphere of perception/action. This phase was time limited to 12 min (avg. 15 animals found).

The HRTFs used were decomposed into minimum phase (for spectral cues) and pure delay (for ITD cues). Thanks to this separation, the subjects could be presented with hybrid HRTFs, where the individual ITD was kept and combined with a non-individual minimum phase component. In order to determine the non-individual component, subjects passed a pre-test where they judged, for each of a 50 HRTF set database, if the perception of a rotating noise burst was "excellent", "good", or "bad". The non-individual HRTF chosen was selected from the "bad" rated while also being highly rated by other subjects (selecting the best of the worse) in order to eliminate the possibility of an HRTF of a non-localizer. 10 subjects (three women, seven men, mean age 28 years) participated in the experiment. None had any known hearing deficit. Subjects were divided into two groups (5 persons each): a test group with non-individual hybrid HRTFs and a control group which used their own individual HRTFs. The control group was used to separate improvements due to a possible "learning effect" of the localization task from true adaptation to the non-individual HRTFs. It is expected that improvements are only to be found in the perception of elevation or front/back confusions and not for lateralization since both groups had their own individual ITD cues.

Results

A first evaluation of localization performance was done in a global manner using spherical angle error. Results were also analyzed concerning three additional errors corresponding to the axes left/right, up/down and front/back as defined by Wightman and Kistler [5]. For front/back errors we observed the confusion rate rather than a mean error angle. Table 1 shows the mean error results obtained with the two localization tests for both groups. It appears that there is a slight improvement in the test group, with localization performance approaching that of the control group after the adaptation phase. The distinction may not be very clear since mean errors are globally quite high (~30° for spherical angle). Nevertheless, as shown in Table 2, there is an observable difference between the two groups progression results (6.1° vs. 0.4° with spherical angle, ANOVA p=0.034) and more precisely this difference is mainly due to elevation results (5.9° vs. 0.1°, p=0,038) and front/back confusion rate (9.3% vs. 2.2%) rather than lateralization (2.8° and 0.7°, p=0.43) where no real modification was observed for either group, as expected. As improvements were more significant for the test group, this suggests an adaptation to the non-individual HRTFs by the auditory system, and not a general training effect for the spatial localization task.

Conclusion

This experiment showed that a rapid adaptation of the adult human auditory system to spectrally non-individual HRTFs may be possible through a spatial map recalibration with multimodal associations. A controlled learning environment was designed with auditory-kinesthetic interaction (with no influence of vision) to elicit adaptation to non-individual spectral cues. Adaptation effects, concerning the perception of elevation and front/back confusions rates, were observed in comparison to a control group. After the adaptation session, localization performance of the two groups converged. These results are interesting in the context of binaural 3D displays faced with the individualization problem and they also suggest that vision is not compulsory to properly calibrate spatial hearing. This is a primarily study using only a few subjects and should be repeated with a larger population to evaluate its robustness. Nonetheless, the results encourage further investigations in the understanding and control of the adaptation process.

Results test1 | test2	spherical angle		up/down angle		left/right angle		front/back conf. rate	
Test group	35	29	27	21	17	14	35	25
Variance	4.2	2.0	4.1	2.2	3.5	2.0	9.2	7.9
Control group	31	30	22	22	16	15	31	29
Variance	5.7	7.1	5.9	6.5	4	4.6	10.8	11.7

Table 1: Mean localization errors before/after adaptation.

Increase in performance	spherical angle	up/down angle	left/right angle	Front/back conf. rate
Test group	6.1	5.9	2.8	9.3
Control group	0.4	0.1	0.7	2.2

Table 2: Progression results defined as the difference between mean errors (= before – after).

References

[1] J. Blauert. Spatial Hearing. MIT Press, Cambridge, 1996.

[2] D.R. Begault, 3-D Sound for Virtual Reality and Multimedia. Academic Press, Cambridge, MA, 1994.

[3] B.F.G. Katz, Boundary element method calculation of individual head-related transfer function. I. Rigid model calculation. J Acoust Soc Am, vol. 110(5): 2240-2248, 2001.

[4] V. Larcher. Techniques de spatialisation des sons pour la réalité virtuelle. PhD Thesis, Université Pierre & Marie Curie, IRCAM, Paris, 2001.

[5] F.L. Wigthman & D.J. Kistler. Headphone simulation of free-field listening. II: Psychophysical validation. J Acoust Soc Am, 85: 868–878, 1989.

[6] F.L. Wightman & D.J. Kistler. Resolution of front-back ambiguity in spatial hearing by listener and source movement. J Acoust Soc Am, 105(5): 2841–2853, 1999.

[7] J.M. Pernaux. Spatialisation du son par les techniques binaurales: application aux services de télécommunications. PhD Thesis, INPG, France Telecom R&D, Lannion, 2003.

[8] P. Hofman, J.A.G. Riswick, & A.J. Van Opstal. Relearning sound localization with new ears. Nature Neuroscience, 1(5), 1998.

[9] P. Zahorik, C. Tam, K. Wang, P. Bangayan, & V. Sundareswaran. Localization accuracy in 3-D sound displays : the role of visual-feedback training. Adv Displays Consortium: ARL's 5th Fed. Lab. Ann. Symp. 17–22, 2001.

[10] G.H. Recanzone. Rapidly induced auditory plasticity : The ventriloquism aftereffect. Proc. Natl. Acad. Sci. v.95, 869–875, Feb. 1998.

[11] I. Viaud-Delmon, L. Sarlat, O. Warusfel. Localization of Auditory Sources in Virtual Reality. Proc. CFA/DAGA, mars 2004.

[12] M.P. Zwiers, A.J. Van Opstal, & J.R.M. Cruysberg. A spatial hearing deficit in early-blind humans. The Journal of Neuroscience, 21: RC142(1–5), 2001.

[13] J. Lewald. Vertical sound localization in blind humans. Neuropsychologia, 40 :1868–1872, 2002.