Rapid head-related transfer function adaptation using a virtual auditory environment

Gaëtan Parseihian[a] and Brian F. G. Katz
LIMSI-CNRS, BP133, Université Paris Sud, Orsay 91403, France

(Received 28 October 2011; revised 25 January 2012; accepted 25 January 2012)

The paper reports on the ability of people to rapidly adapt in localizing virtual sound sources in both azimuth and elevation when listening to sounds synthesized using non-individualized head-related transfer functions (HRTFs). Participants were placed within an audio-kinesthetic Virtual Auditory Environment (VAE) platform that allows association of the physical position of a virtual sound source with an alternate set of acoustic spectral cues through the use of a tracked physical ball manipulated by the subject. This set-up offers a natural perception-action coupling, which is not limited to the visual field of view. The experiment consisted of three sessions: an initial localization test to evaluate participants' performance, an adaptation session, and a subsequent localization test. A reference control group was included using individual measured HRTFs. Results show significant improvement in localization performance. Relative to the control group, participants using non-individual HRTFs reduced localization errors in elevation by 10° with three sessions of 12 min. No significant improvement was found for azimuthal errors or for single session adaptation.
© *2012 Acoustical Society of America*. [http://dx.doi.org/10.1121/1.3687448]

PACS number(s): 43.66.Pn, 43.66.Qp, 43.66.Lj [NAG] Pages: 2948–2957

I. INTRODUCTION

The human auditory system decodes sound source location using a set of acoustic cues that are contained in the head-related transfer function (HRTF).[1] The localization cues can be divided into binaural disparity cues, principally linked with the lateralization of sound sources, and spectral cues, which are linked with elevation discrimination. The procedure for creating virtual audio 3D sound using binaural synthesis consists of processing an audio signal with the HRTF for a given position and then presenting it using stereo headphones.[2] The HRTF varies according to morphological factors (shape of outer ears, head dimensions, torso) thus providing individual cues. It is well known that using non-individual HRTF in virtual auditory environments (VAE) results in perceptual distortions such as front/back confusions, angular distortions in the vertical plane, and non-externalized auditory images.[3]

The impossibility of measuring and employing the individual HRTF for each and every potential listener in a commercial use situation has prompted many studies on the problem of individualization. Most of these studies focus on scaling the HRTF to the individual using morphological criteria,[4,5] tuning of spectral cues,[6] numerical computation[7,8] or subjective selection.[9,10] While some proposals exist for the adjustment of time difference cues to the individual,[11] the rapid individualization of the spectral components still poses computational and technical problems.

Binaural synthesis processed using individual HRTF is considered to provide high fidelity rendering of the auditory scene, but even in the best conditions some artifacts may remain due to the unnaturalness of the VAE. In order to minimize these types of anomalies and to habituate the subject to the VAE, studies using binaural spatialization often use trial sessions.[12–14] This effect, often termed the "learning effect" is a procedural learning that regroups the performance improvement due to familiarization with the task, the stimuli, and the report method. In the present study, this effect is used to consider the individualization problem in the reverse sense. Instead of adapting the HRTF to the individual, this study attempts to force the auditory system to quickly adapt to a non-individual HRTF.

Naturally, the mammalian auditory system is able to make use of sensory experience to calibrate certain aspects during early experience and to make limited adjustments during development in order to maintain accuracy and precision.[15] Previous studies have demonstrated the ability of animal auditory systems to calibrate and adapt to major changes in frequency, temporal sequence, sound level, and sound localization.[16] The first study highlighting the plasticity of human sound localization system, by Young,[17] presents a diary of the subject/author who wore a "pseudo-phone" for a period of 18 consecutive days, describing his feelings and any changes in spatial auditory perception. The pseudo-phone is an instrument for producing illusory auditory localization by changing the relationship between the ears and the actual direction of the sound. It consists of two ear trumpets worn on the head, each connected to the ear canals on the opposite side of the head via a sound-proof tube. The result is twofold, an inversion of the ITD and a simplification of the HRTF spectral cues due to the shape of the trumpet replacing the pinnae. Young wore this device 1 hour daily for the first 9 days, 2 hours daily for the following 6 days, and continuously for 3 complete days. During the first days, he reported a dissociation of visual and auditory localizations, right-left reversals, vague and double localizations, and unlocalizable sounds. He noticed changes from reversed

[a]Author to whom correspondence should be addressed. Electronic mail: gaetan.parseihian@limsi.fr

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

to normal localization when the position of the source became known, typically when entering the visual field of view. After 18 days, there was no habituation to auditory localization with eyes closed, whereas with eyes open all of the sounds in a complex situation were normally localized. Upon removal of the pseudo-phone, localization perception immediately returned to normal and no subsequent disturbances in localization were noted. A study by Hofman *et al.*[18] on four subjects, revealed that people with "modified pinnae," which altered their spectral elevation cues, steadily reacquired localization abilities in the visual field of view after several weeks of passive adaptation. Moreover, it appeared that the auditory system may retain the capacity to simultaneously decode multiple sets of spatial cues, individual and modified (with pinnae mold inserts) as once the molds were removed, subjects quickly found their bearings, and returned to their previous performance ability. It should be noted that the use of inserts can only reduce geometrical elements of the pinnae (e.g., the concha cannot be made larger, only smaller), and so always shift resonant properties of HRTF to higher frequencies. As such, this study highlighted an adaptation of individuals to smaller ears, but does not examine the case of bigger ears or frequency lowering of spectral cues. In a cross modal study, Zwiers *et al.*[19] showed that spatially scaled vision (using compressing 0.5× lenses for nine subjects) induced systematic and adaptive changes in sound localization that restore the spatial calibration between the two modalities within 2–3 days of adaptation. These changes were a compression of auditory localization, which was most pronounced for azimuth and mainly restricted to the visual field of the lenses. Outside the field of visual-auditory interaction, sound localization was also affected, but in contrast was expanded toward the visual field.

Studies on blind individuals have allowed researchers to further investigate the role of vision and other senses on the adaptation of the auditory system. In an attempt to demonstrate that visual feedback might be essential for a full development of human sound localization, Zwiers *et al.*[20] showed a spatial hearing deficit of blind people in the frontal hemisphere. However, other studies[21,22] have demonstrated that vision is not essential and that other sensory input (e.g., tactile and motor feedback) might be sufficient for the development of sound localization in both azimuth and elevation. Lessard *et al.*[22] demonstrated that blind subjects can localize binaurally presented sounds as well as sighted individuals in the azimuthal plane. Lewald[23] found poorer performance for blind subjects in vertical localization but posed the problem of the point of reference to extrapolate the indicated position when using a pointing reporting protocol. His results indicate that blind subject's point of reference to indicate a source position is not the center of the head. This result was confirmed by Zwiers *et al.*[24] who found equal performance when taking into account a modified point of reference between sighted and blind subjects, being either head or shoulder based. Although vision is consider the main sense allowing the calibration of sound localization, these results tend to demonstrate that other sensory modalities also play essential roles in sound localization calibration.

In the context of virtual audio environments, the use of a 3D sound display synthesized with non-individual HRTFs may be considered as "listening with the ears of someone else." Previous studies have demonstrated that non-individual virtual displays produce distorted spectral cues and non-adapted binaural cues,[3] inducing the same effect of the pinna-mold manipulation used by Hofman *et al.*[18] but in a virtual situation. One difference between these two types of modifications is that pinna insert molds[18] or compressing lenses[19] can be worn permanently for a few days while it is not feasible to continually use a VAE in everyday life. Conversely, modifications can be made which are not possible via physical modifications.

Studies by Shinn-Cunningham *et al.*[25,26] using visual feedback have shown that, in several sessions of 2 hours, listeners can adapt to virtually distorted horizontal localization cues. "Supernormal" cues were created by enlarging the cues of a given HRTF in the front region and reducing those corresponding to peripheral locations. With this transformation, a source from azimuth θ was synthesized using the HRTF that normally corresponded to position $f(\theta)$. Three to eight subjects completed several localization and adaptation tasks with the same HRTF set. The spatial resolution was initially enhanced for the frontal regions and degraded for peripheral locations. With repeated training, subjects appeared to adapt to the new relationship between the acoustic cues and positions in space leading to a reduction of the minimum audible angle in the lateral dimension. Zahorik *et al.*[27] highlighted a rapid adaptation (two sessions of 30 min) to spectral indices of an HRTF using visual feedback. Twelve subjects listened to binaurally synthesized stimuli in an immersive audio-visual training platform composed of a Head Mounted Display and a head-tracking device. The learning task was similar to a localization test, where the subject must first point their nose in the direction of the perceived virtual source, which was presented once. Feedback was provided afterwards by presenting a bright light at the virtual sound source's position, while the stimulus was repeated. The subject was then required to point their nose towards the bright spot, corresponding to the correct position, before continuing to the next stimulus. Participants repeated this task two times over four days. Results showed an improvement in front/back discrimination where the proportion of hemifield reversed responses decreased from 38% to 23%. In contrast, there was no observable improvement in elevation discrimination performance. It was noted that the improvements appeared to be retained for at least four months. Combining real and virtual learning, Honda *et al.*[28] explored the effects of playing a virtual auditory game on performance in a real sound localization task. The game consisted of hitting a buzzing virtual honeybee with a tracked plastic hammer. Participants were equipped with head tracked headphones and the VAE was rendered with respect to head movements. Real sound localization performance was evaluated using an array of 36 loudspeakers positioned at 30° intervals at elevations 0° and ±30° with a radius of 1.2 m. Pointing responses were interpreted by an operator as being the closest point on the loudspeaker grid. With seven sessions of 30 min of this perceptual-motor training using virtual audio and non-visual feedback, results revealed that localization performance for real sounds was significantly improved.

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

The current study focuses on perceptual adaptation to the spectral component of the HRTF over the entire sphere in the context of virtual rendering with rapid training, inspired by a preliminary study conducted on 10 subjects by Blum *et al.*[29] The aim is to quantify the adaptation effect using a quick exploration of the spatial map by an auditory-kinesthetic process. Each subject was placed in a VAE and directly controlled the position of a virtual sound through the use of a hand-held tracker. The principle is that the listener associates the source position with the dynamic acoustic cues used for binaural rendering through the constant and innate awareness of one's own hand position. This study explores the effect of this multi-modal training platform taking into account both the similarity of the individual's HRTF to the training HRTF and the number of training sessions. The goal is to determine the relative performance of a proprioceptive/vestibular feedback in a training procedure of sound localization. The results are compared to a control group who used their individual HRTF in order to separate procedural learning (where the participants merely become familiar with spatial distortions introduced by the system) from perceptual learning (where the participants' actual perceptions change).

II. METHODS

A. Subjects

A total of 24 adult subjects (5 women, 19 men, age between 20 and 60 years) served as paid volunteers; none had any known hearing deficit. All were naive regarding the purpose of the experiment and the sets of spatial positions selected for the experiment. Three were familiar with VAE and with sound localization studies, and only one was experienced with the use of his own HRTFs in VAE. The subjects who took part in this study were the same as those who took part in the associated previous study by Katz and Parseihian.[31]

B. Non-individual HRTF selection

The HRTFs used were comprised of a set of 7 HRTFs selected from the 46 HRTFs of the public database LISTEN[30] using the method described in Katz and Parseihian.[31] The HRTF were decomposed into spectral component (representing spectral cues) and pure delay (representing ITD cues). Participant used a hybrid HRTF, where the modeled individual interaural time difference (ITD) based on head circumference was combined with a selected spectral component. A VAE quality classification test was used in order to rate the different HRTFs for each subject, using a continuous scale with extremes being "good" and "bad" by evaluating the rendering of two predefined source trajectories. The first trajectory was a circle in the horizontal plane with points at 30° spacing. The trajectory began directly left and followed two complete rotations around the subject. The second trajectory followed an arc in the median plane (azimuth = 0°) from elevation −45° in front to −45° at the rear with points at 15° spacing. The trajectory commenced in front, proceeded to the top and to the rear, and then returned along the same path to the front. The stimuli source was a repeated noise burst, 0.23 s in duration and shaped with a Hann function window. Subjects had to judge each of the seven HRTFs sets on the scale from "good" to "bad" for each trajectory. An overall judgment rating was taken as the sum of the two trajectory judgments. This rating served in selecting each subject's non-individual HRTFs in the experiment.

C. Design and procedure

To assess the effects of the training sessions on sound localization performance, two different tasks were repeated several times over the course of three days. An *adaptation session* (*A*), designed like a game, where blindfolded subjects had total control of a virtual sound source spatialized at their hand position, permitted a rapid training using the natural interactivity with perception/action coupling. A classical *localization test* (*L*) allowed the evaluation of the performance of the subject during each phase of the experiment. In order to evaluate the effect of the similarity between the subject's HRTF and the non-individual HRTF on the adaptation, subjects began the experiment by a *selection test* which consisted of perceptually classifying several HRTFs sets. For details of the HRTF database and selection test, see Katz and Parseihian.[31] Following this, the 20 subjects using non-individual HRTF were randomly divided into two groups; half were assigned their highest rated HRTF (group *good, G*) taken from the results of the selection test (see Sec. II B) and for the entire experiment they heard sounds synthesized by using this selected HRTF which was quite similar to their own HRTF. The other half of the subjects were assigned their lowest rated HRTF (group *bad, B*) from the selection test (see Sec. II B) which meant that they heard sounds synthesized by using this selected HRTF which was quite different to their own HRTF). Four subjects with their individual HRTFs available were included as a control group (*C*). To evaluate the effect of time and repetition with regards to HRTF adaptation, the experiment was performed over three consecutive days. For each of the two test groups (*G* and *B*), half performed one *adaptation task* (*A*) just the first day, while the other half performed the *adaptation task* once each day (i.e., three times). The control group performed the *adaptation task* once each day. Each subject performed a total of four *localization tests* (L_1, L_2, L_3, and L_4); once prior to the first adaptation session to evaluate their initial performance and then after each adaptation session for the second set of groups (*C*3, *G*3, and *B*3) or one per day for the other groups (*G*1 and *B*1). Table I summarizes the configuration of the experiment for each group.

TABLE I. Configuration of the groups of participants.

Group type	Nb	HRTF type	Day 1	Day 2	Day 3
*G*1	5	*Good select.*	$L1 \rightarrow A \rightarrow L2$	$L3$	$L4$
*G*3	5	*Good select.*	$L1 \rightarrow A \rightarrow L2$	$A \rightarrow L3$	$A \rightarrow L4$
*B*1	5	*Bad select.*	$L1 \rightarrow A \rightarrow L2$	$L3$	$L4$
*B*3	5	*Bad select.*	$L1 \rightarrow A \rightarrow L2$	$A \rightarrow L3$	$A \rightarrow L4$
*C*3	4	Individual	$L1 \rightarrow A \rightarrow L2$	$A \rightarrow L3$	$A \rightarrow L4$

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

For the entire experiment, binaural sound sources were rendered using the LIMSI Spatialisation Engine,[32] a real-time spatialization engine in Max/MSP based on full-phase HRIR convolution in contrast to the minimum phase HRIR approach often used). The subjects were equipped with a stereo open ear headphone (model Sennheiser HD570) tracked with a 6-DoF position/orientation magnetic sensor positioned on the top of the headphone. They held a position-tracked ball in their hand and interacted with the system using a foot pedal. The position of the hand was calculated relative to the 6-DoF head tracker shifted to the center of the head for the sound rendering in the *adaptation task* and the pointing task in the *localization test*. All phases of the experiment were conducted in the same quiet room (35 dBA SPL background noise level) with the listener seated in a swivel chair with their foot near the response pedal. No headphone equalization was used, in order to present a situation comparable to a real application case where a novice subject uses his own headphones.

D. Adaptation task

The concept of the *adaptation task* (*A*) was to immerse the user in a virtual audio environment that included a proprioceptive feedback of the auditory spatial information. In order to perform this training subconsciously, the task was formulated as a game-like scenario, and as such the subject remained unaware of the aim of the experiment. The game consisted of the subject searching for animal sounds hidden around him/her, scanning the space with the hand-held position-tracked ball. The search feedback sound consisted of alternating pink/white noise (50–20 000 Hz) with an overall level of approximately 55 dBA measured at the ear. The delay between the bursts decreased as the angular distance between the ball and the hidden target position reduced (following a Geiger counter or sonar metaphor), from 3.0 s to 0.2 s, with a 5 ms onset/offset hamming ramp. Full spectrum burst sounds were chosen in order to favor an adaptation to the complete spectral cues of the HRTF. The "detector sound" was spatialized at the ball center, with the virtual rendering position updated every 50 ms with respect to listener's head position/orientation and the ball position. When the target position was found, the detector sound was replaced by a random "animal sound," still rendered at the ball position, that the subject could then move about for several seconds. The animal sounds were taken from various free sample databases; their mean duration was 5 ± 2 s. The position of the hidden target was selected randomly from a list of 50 HRTF measurement positions in order to ensure the subject explored the entire auditory sphere. Subjects were instructed to perform the game task as entertainment. They were free to move with the swivel chair in order to facilitate reaching rear positions, but they had to keep their foot near the response pedal. The duration of the adaptation task was fixed to 12 min so that subjects had time to find more than the half of the targets and that they could explore the entire sphere.

E. Localization task

The *localization task* (*L*) consisted of reporting the perceived position of a static spatialized sound sample using a hand pointing technique validated by a foot pedal. This judgment elicitation technique and method using a body part has the ecological advantage of being egocentric and natural for the user and is the best choice for experimental measurement of 3D positions by blind subjects.[33] Each subject was instructed to orient him or herself straight ahead and to keep their head fixed during the short sound stimuli presentation. The stimulus was short to exclude head movement effects. It consisted of a train of three, 40 ms Gaussian broadband noise bursts (50–20 000 Hz) with 2 ms, Hamming ramps at onset and offset and 30 ms of silence between each burst. This stimulus was chosen following the study of Dramas *et al.*[34] where the effect of repetition and duration of the burst on localization accuracy was analyzed. Their results showed an improvement of the accuracy between three repeated 40 ms bursts and a single 200 ms burst. The overall level of the train was approximately 55 dBA measured at the ears.

After presentation of the stimulus, each subject was instructed to point his hand (ball in hand) in the direction of the perceived sound source location and to validate the response with the foot pedal. For holding the ball, no hand was imposed, and the hand holding the ball could be changed at will. The perceived orientation was calculated between the initial head position/orientation when the stimulus was played and the final hand position when the listener validated the target. No feedback was given to the subject regarding the target position.

A total of 25 positions (see Table II) were randomly presented with 5 repetitions each. This partial sphere included a full 360° of azimuth, and −40° to 90° of elevation relative to ear level. Subjects had to localize a total of 125 targets and were naive with respect to the set of spatial positions selected for the experiment. The mean duration of this task was 10 min.

F. Spatial coordinates and data analysis

The analysis of localization performance, which was originally recorded in standard spherical coordinates (azimuth and elevation), was performed using the interaural polar coordinate system (see Morimoto *et al.*[35] In this coordinate system, azimuth and elevation angles are transformed into lateral and polar (or rising) angles, with the polar angle rotation axis

TABLE II. Lateral (θ) and polar (ϕ) angle of the 25 positions used for the localization test (in interaural polar coordinate system, see Sec. II F).

Median		Front		Back	
Lateral	Polar	Lateral	Polar	Lateral	Polar
0	0	30	0	−30	180
0	−30	90	0	30	180
0	30	−90	0	−23	−158
0	90	−30	0	−8	140
0	150	23	22	44	166
0	180	8	−40	11	−140
0	210	−44	−14	−26	106
		−11	41	10	107
		27	74		
		−10	73		

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

being the interaural axis. The direction of a vector between the head center and a point on the sphere is expressed by two angles: the lateral angle and the polar angle. The angle between the vector and the median plane is the lateral angle, from $-90°$ to $90°$. The polar angle corresponds to rotation around the interaural axis, from $-90°$ to $270°$, with $0°$ being directly in front. This is a natural coordinate system for human localization data since it allows for the rough separation of temporal cues, which are related to the ITD and are represented by the lateral angle, from the spectral cues, which are related to the HRTF and are represented by the polar angle. Using the interaural polar coordinate system, all front/back and up/down confusion errors are contained in the polar angle. Localization errors in lateral and polar angles were analyzed by regarding the magnitude of the difference between the target angle and the perceived angle.

III. RESULTS

A. Localization test

1. Lateral angle error

Summary results for lateral angle error are shown in Fig. 1, indicating the evolution of mean error for each group and for each localization test. Overall localization blur was from $13°$ to $16°$ for the *control* group and from $17°$ to $25°$ for the other groups. The *control* group ($C3$) exhibited better performance relative to the other subject groups, which can be explained by approximations employed in generating the individualized ITD model used for the hybrid non-individual HRTFs. Since the ITD model was not perfect, some improvement on lateral performance after the adaptation session could be possible. An improvement was found for group $G3$, while no difference over the course of the four tests was seen for groups $G1$, $B1$, or $B3$. An analysis of variance (ANOVA) test showed that the difference of $5°$ between $L1$ and $L4$ for group $G3$ was marginally significant [$F_{1,8} = 4.82$, $p = 0.06$] and this improvement indicates a tendency of adaptation to the non-individual ITD cues. An analysis of the effect of the type of HRTF for the first test $L1$ (combining results of the two groups with *good* HRTF ($G1$ and $G3$) into a single group G and the results of the two group with *bad* HRTF ($B1$ and $B3$) into a single group B highlights a significant difference between the group $C3$ and groups G [$F_{1,12} = 16.01$, $p < 0.005$] and between $C3$ and B [$F_{1,12} = 15.75$, $p < 0.005$]. No significant difference was found between groups G and B [$F_{1,18} = 2.25$, $p = 0.15$)] for $L1$.

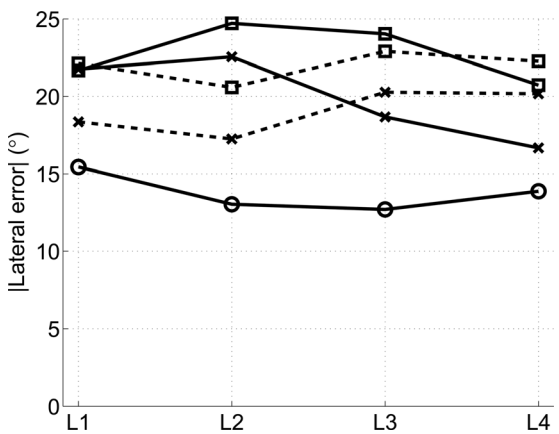

FIG. 1. Mean magnitude lateral error for every group: (×) *Good,* (□) *Bad,* (○) *Control;* (…) for $G1$ and $B1$ (one adaptation session), (–) for $G3$, $B3$ and $C3$ (three adaptation sessions).

2. Polar angle error

Figure 2 shows the results for the polar angle error with a representation combining boxplot, histogram, and the mean magnitude error for each group and each localization test.

This type of representation has the advantage of combining a boxplot (left side) containing traditional statistical data [lower quartile (Q1), median (Q2), upper quartile (Q3)] with a histogram (right side), representing the distribution of the response errors. As the polar angle contains all front/back and up/down confusions, no resolution or suppression of these types of errors was performed in order to observe their evolution on the distribution of the responses. For the first test $L1$, the distribution for the control group $C3$ was normal while the distribution of polar error for the groups with non-individual HRTFs ($G1$, $G3$, $B1$, and $B3$) was multimodal and highlighted many confusions error. As the major part of the histograms highlight a polar error which is not uniformly distributed, a traditional ANOVA cannot be performed. Instead, a Kruskal-Wallis test was performed.

For the first localization test, the mean errors were approximately $55°$ for the control group and between $70°$ and $80°$ for the other groups. A Kruskal-Wallis test on the mean error for all positions and repetitions for each subject between only $L1$ and $L2$, combining groups ($G1 + G3 \rightarrow G$, $B1 + B3 \rightarrow B$), demonstrated a significant difference between $C3$ and G [$\chi^2_{3,24} = 17.98$, $p<0.001$]. *Post hoc* analyses (Tukey's) revealed significant differences between $C3$ and G for $L1$ ($p < 0.001$) and $L2$ ($p < 0.005$). Differences between $C3$ and B were also observed [$\chi^2_{3,24} = 17.02$, $p<0.001$]; *post hoc* analyses (Tukey's) revealed significant differences between $C3$ and B for $L1$ ($p < 0.005$) and $L2$ ($p < 0.005$). Smaller but significant differences were observed comparing groups G and B [$\chi^2_{3,36} = 16.58$, $p<0.001$] with $p < 0.05$ for $L1$ and $p < 0.01$ for $L2$.

The evolution of error values over the four localization tests indicated a significant improvement in performance for the groups which performed three adaptation sessions ($C3$, $G3$, and $B3$) and only a slight to moderate improvement for the two groups which performed only one adaptation session ($G1$ and $B1$). For the *control* group, most improvement occurred after the first adaptation session ($L1 \rightarrow L2$) with an improvement of $10°$, and only a few degrees afterwards ($L2 \rightarrow L4$). A Kruskal-Wallis test showed significant difference between $L1$ and $L2\&L3\&L4$ [$\chi^2_{3,12} = 8.93, p<0.05$]. Evolution of performance for group $G3$ was constant throughout the test (improvement of $17°$ on the mean and $23°$ on the median overall) with a significant difference between $L1$ and $L4$ [$\chi^3_{3,16} = 11.48, p<0.01$], while group $B3$ improvement appeared principally after the second learning session ($L2 \rightarrow L4$) with an overall improvement of $15°$ on the mean and $24°$ on the median error. A Kruskal-Wallis test showed

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

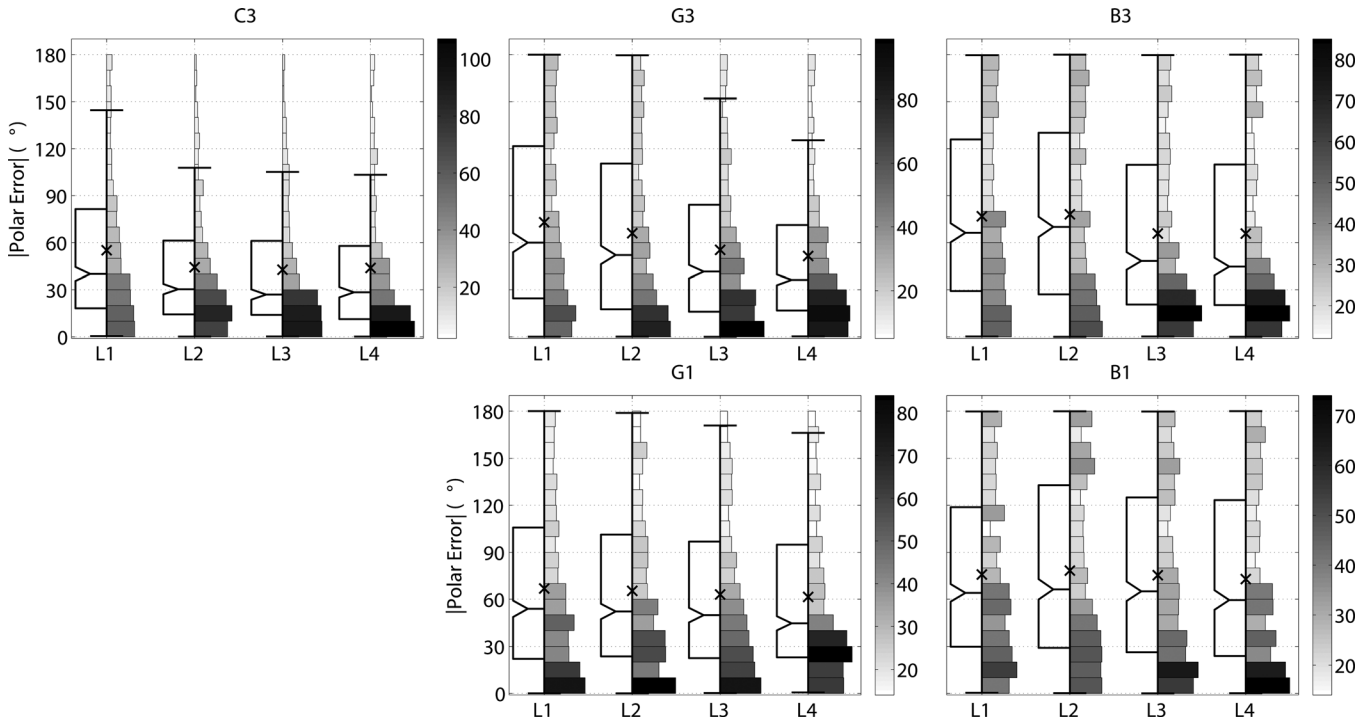

FIG. 2. Split boxplot-histogram of the magnitude of polar error by test for group *C*3 (top-left), *G*3 (top-middle), *B*3 (top-right), *G*1 (bottom-middle) and *B*1 (bottom-right). Boxplot and angular error scale at the left; mean values (×); histogram value legend color bar on the right.

significant difference for group *B*3 between *L*1&*L*2 and *L*4 [$\chi^2_{3,16} = 15.25$, $p<0.005$]. This is most evident through an inspection of the histogram of response errors, as shown in Fig. 2. The observation of the error distribution for groups *G*3 and *B*3 showed a transformation from multimodal distribution (in *L*1) to normal distribution (in *L*4) for *G*3 and a compression near 15° of error for *B*3. It should be noted that the performances of group *G*3 for the final test *L*4 was comparable to the initial performances of the control group, *C*3:*L*1. A Kruskal-Wallis test between *G*3:*L*4 and *C*3:*L*1 showed no differences [$\chi^2_{1,7} = 0.96$]. Furthermore, final performances of group *B*3 were quasi-comparable to the performance of group *G*3 during test *L*2 [$\chi^2_{1,8} = 0.32$].

A slight evolution of performance was observed for the two groups with only one adaptation session (*G*1 and *B*1), where a significant difference between tests *L*3 and *L*4 (due to a concentration of polar error near 0°) was found for group *B*1 [$\chi^2_{3,16} = 11.07$, $p<0.05$] and no significant difference was found for the group *G*1 [$\chi^2_{3,16} = 5.03$, $p<0.17$]. During the four localization sessions, there was an overall improvement of 8° on the mean (9° on the median) for group *B*1 and 7° on the mean (10° on the median) for group *G*1, comparable to the learning effect seen with the control group of 10°.

A linear regression analysis was performed on the polar angle responses. The mean and standard deviation across subjects of the slope of the regression line and goodness-of-fit criteria r^2 for each group and each test are shown in Table III. Since no correction or suppression of confusion errors were applied on the data, regression slope lines were far from the unity expected for a perfect localization of virtual sound. The results also highlighted a large inter-subject variability with a large standard deviation. Analysis of the effect of the adaptation task on the slope of the regression line showed a difference in improvement between the two groups with three training sessions (*G*3 and *B*3) and the other group (*G*1, *B*1 and *C*3). In effect, between the first and the last localization test, the slope of regression line increased by a factor of 3.27 for *G*3 and 3.13 for *B*3 whereas it only increased by a factor of 1.05 for *G*1, 1.46 for *B*1, and 1.15 for *C*3. It can be noticed that the regression slope lines of control group (between 0.52 and 0.61) are closer to unity than regression slope lines of the other group (between 0.07 and 0.43), highlighting the difference between subjects with individual HRTFs and subjects with non-individual HRTFs.

In summary, a significant difference between the three groups was observed before the first training session

TABLE III. Mean linear regression analysis and goodness-of-fit criteria r^2. Variances shown in parenthesis.

	Regression slope					r^2				
Test	*C*3	*G*1	*G*3	*B*1	*B*3	*C*3	*G*1	*G*3	*B*1	*B*3
L1	0.52 (.22)	0.28 (.23)	0.13 (.10)	0.08 (.03)	0.07 (.13)	0.24 (.14)	0.13 (.16)	0.05 (.06)	0.03 (.02)	0.02 (.03)
L2	0.60 (.24)	0.20 (.22)	0.12 (.17)	0.05 (.04)	0.08 (.11)	0.44 (.35)	0.10 (.13)	0.09 (.14)	0.03 (.05)	0.02 (.02)
L3	0.61 (.24)	0.35 (.21)	0.27 (.22)	0.06 (.06)	0.28 (.20)	0.40 (.27)	0.16 (.18)	0.18 (.17)	0.02 (.03)	0.11 (.17)
L4	0.60 (.29)	0.30 (.18)	0.43 (.12)	0.11 (.13)	0.23 (.18)	0.38 (.30)	0.12 (.11)	0.27 (.14)	0.05 (.09)	0.10 (.11)

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

highlighting the degree of degradation induced by the use of nonindividual HRTF, with better performances for the control group and worst performances for the group who were assigned their lowest rated HRTF. In terms of localization improvement, the two groups with three adaptation sessions (*G*3 and *B*3) significantly enhanced their performances and reduced their median error by approximately 10° relative to the control group (*C*3). The two groups with only one adaptation session (*G*1 and *B*1) enhanced their performance somewhat, but this improvement was not superior in comparison with to the control group (*C*3) indicating that this can be considered rather as a procedural adaption to the test protocol.

3. Error type

Previous studies, such as Zahorik *et al.*,[27] have quantified adaptation effects by analyzing the change in front/back reversals before and after training sessions. A similar approach has been employed here, with an analysis by error type in more detail, as simple front/back type confusion analysis is not truly appropriate in a full sphere localization task. With the conventional definition of front/back and up/down confusions (proposed by Wightman *et al.*[13] and Wenzel *et al.*[36]), if the angle between the target and the judged position is bigger than the angle between the target and the mirror of the judgment about the vertical plane passing through the subject's ears (or, the horizontal plane passing through the ears), the judgment is considered as a confusion. Another definition, proposed by Martin *et al.*[37] defines front/back confusion according to two conditions. The first is that both the azimuth of the target and the judged position do not fall within a narrow exclusion zone of 15° (±7.5°), at 0° elevation and equal to 15° divided by the elevation's cosine for all elevations, symmetrical about the vertical plane dividing the front and back hemispheres. The second is that the target and the perceived positions are in different front/back hemispheres. This definition is extended to up/down confusions with the same two conditions on the elevation angle by adding another exclusion zone of 15° around the horizontal plane. This results in a response space divided in four zones: one for no confusions (called *precision*), one for front/back confusions (called *front*/*back*), one for up/down confusions (called *up*/*down*) and one combining front/back and up/down confusions (called *combined*). In the interauralpolar coordinate system, all of these zones are contained in the polar angle. Figure 3(a) shows a representation of these zones in polar angle as a function of the target angle. With this definition, if a target position at 0° in lateral and 8° in polar angle is perceived at −8° polar angle, it is marked as an up/down confusion whereas it is clearly in the localization blur. In order to solve this problem, this study purposes a new definition of the errors zones that reorganizes the zones previously proposed to reduce the artifacts near quadrant boundaries, based on a proposal by Yamagishi and Ozawa[38] which is extended here to deal with multiple confusion error types. Figure 3(b) shows error zones that are defined according to a region around an error axis, defined here as a zone of ±45°. The value of ±45° was found empirically by

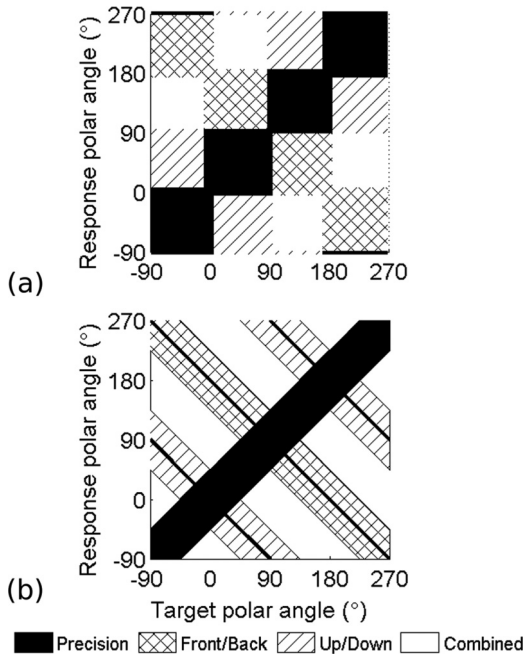

FIG. 3. Definition of the four different error type zones according to (a) Martin *et al.* (Ref. 37) and (b) the proposed zone definitions for clearer treatment of quadrant boundaries.

examining histograms of magnitudes of polar errors across all listeners and conditions inspired by the method used by Middlebrooks.[5] Errors along the principal axis are considered *precision* type errors, while errors outside of both the *front*/*back* and *up*/*down* regions are considered *combined* errors, as they cannot be attributed solely to either *front*/*back* or *up*/*down* confusion errors. A comparison of Fig. 3(a) and Fig. 3(b) shows a sub-estimation of combined errors in the method of Martin *et al.*,[37] and an overestimation of the other types of error. Since this study aims to quantify the effect of the training procedure in term of improvement of precision error rate, the method proposed here seems to be more adequate.

Figure 4 presents the results for polar angle responses for three representative subjects from groups *G*3, *B*3, and *C*3. The results of the *C*3 subject are quite accurate, with few front/back confusions for *L*1 and some errors in the median plane. Front/back errors almost totally disappeared after the first learning session while median plane localization errors remain. The results for the *B*3 subject (with *bad* match non-individual HRTF) are located mostly in the rear hemisphere (polar angle between 90° and 270°) for the localization test *L*1. During the course of the three training sessions, this subject gradually reacquired the capacity to localize some sources in the frontal region. The results for the *G*3 subject (with *good* match non-individual HRTF) were rather poor for *L*1; the evolution of these results highlight the benefit of the three training tasks, with final errors being mostly for elevations below the horizon.

Results of the error type analysis are provided in Table IV with the error distribution by type for each group and each localization test. Ideally, there would be no confusions and all errors would be of the *precision* type. For the control group *C*3, all improvement occurred after the first

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

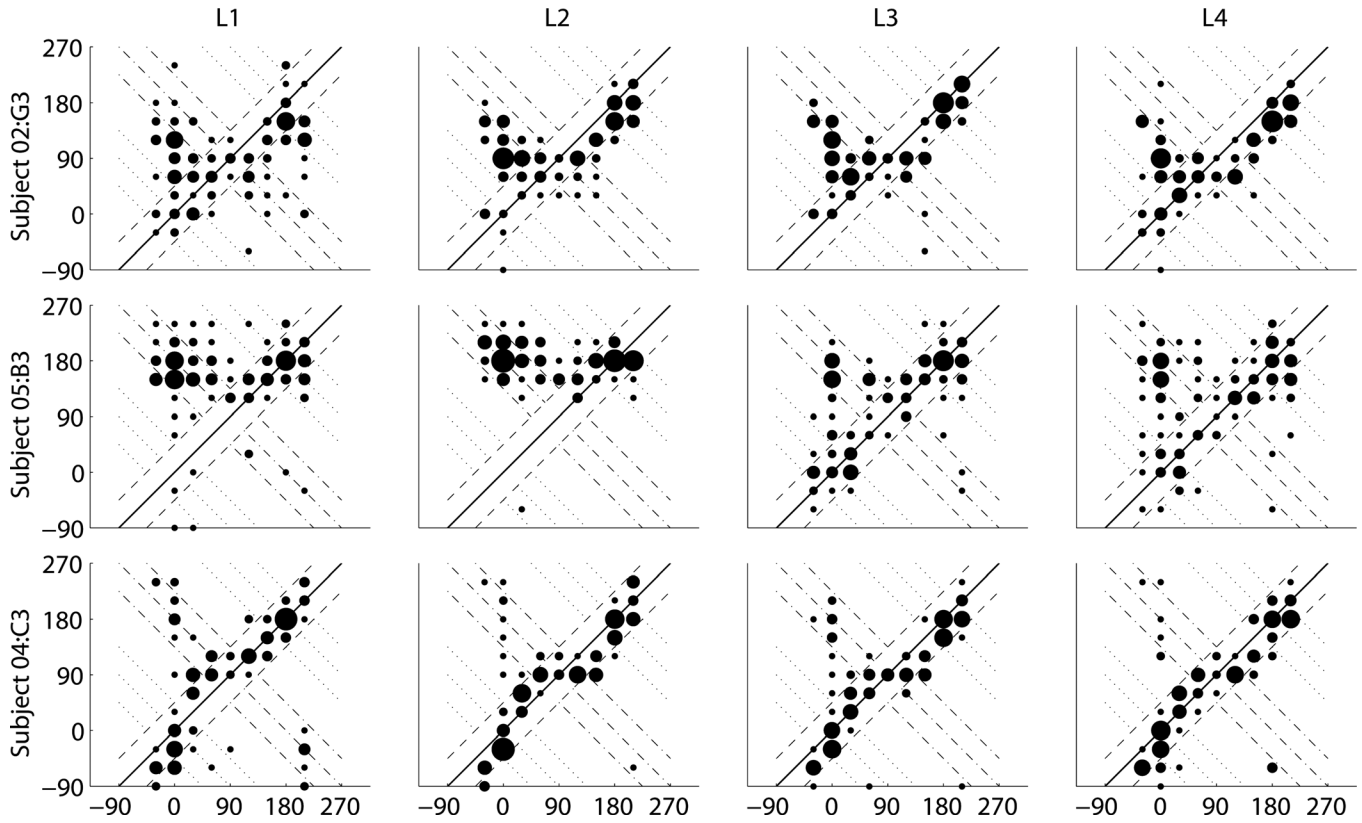

FIG. 4. Evolution of polar angle response for representative subjects of groups *G*3, *B*3, and *C*3. The data are displayed in interaural polar coordinate system for each phase of the experiment (indicated in the column heading).

learning session. The percentage of precision error for *L*1 was 54%, increasing to 66% for *L*2 and remained stable (67%). Overall, the other types of errors decreased from 14% to 12% for front/back confusions, from 7% to 5% for up/down confusions, and from 25% to 17% for combined errors. The increase in *precision* type was therefore principally due to a decrease in *combined* errors. While group *C*3 had a proportion of *precision* errors between 54% and 67%, the other groups (using non-individual HRTF) had a *precision* proportion of approximately 40% for the first test and a maximum of 57% for the final test. The *precision* error rate for group *G*1 increased by 7% between the first and last session. This improvement was completely matched by a reduction in *combined* errors (from 32% to 25%), and no effect on front/back and up/down confusion rates, which remained stable at 18% and 6% respectively. Performance of group *G*3 increased from 41% to 57% for *precision* error rates, which was principally due to front/back confusion reduction (from 25% for *L*1 to 11% for *L*4), with other error types remaining relatively stable. Group *B*1 had practically no evolution of the different types of errors while group *B*3 exhibited an increase in *precision* error rates (from 36% to 50%) reflected by a reduction in all other error types (27% to 23% for front/back confusion, 11% to 6% for up/down confusion, and 26% to 20% for *combined* error).

B. Game task analysis

As the purpose of the adaptation task was not to precisely locate positions in space but just to create a game-like interface for providing audio-kinesthetic feedback to adapt subjects to the non-individual HRTF cues, no detailed data was retained during this phase. Nevertheless, the total number of animals found in each section provides some indication about the difficulty performing the task and its dependency to the type of HRTF (*C, G, B*) or the progress between each session for groups *G*3, *B*3, and *C*3. The results showed no significant effect of HRTF type on the average

TABLE IV. Distribution of error type (percentage) by group.

	C3				G3				B3				G1				B1			
Error Type	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4	L1	L2	L3	L4
Precision	54	66	67	67	41	46	53	57	36	37	48	50	43	45	46	50	40	41	39	43
Front/back	14	10	12	12	25	24	16	11	27	24	21	23	19	18	18	18	24	29	28	28
Up/down	7	7	5	5	8	5	7	7	11	10	9	6	6	6	7	6	11	7	11	8
Combined	25	17	16	17	27	25	25	25	26	28	23	20	32	31	29	25	25	23	22	21

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

number of animals found with subjects finding a mean of 28.4 ± 6.1 animals in the first session of 12 min. An improvement of approximately 6 more animals was seen for the second session (mean of 34.7 ± 4.8 animals found) but this difference was not significant. Moreover, there was no improvement between the second and third sessions (mean of 35.1 ± 6.4 animals found). This same evolution for the three groups could be explained by a familiarization to the task more than an effect of adaptation to the HRTF cues. Moreover, subjects reported employing different strategies in performing the task: some tried to find the most animals possible while others spent more time experimenting with the sound positioned in their hand, thereby finding fewer animals in the 12 min time limit.

IV. DISCUSSION

The purpose of this study was to explore the effect of an audio-kinesthetic environment on rapid auditory map recalibration. An evaluation of the plasticity of the human auditory system using a rapid audio-kinesthetic virtual auditory environment demonstrated that at least two adaptation sessions were required in order to obtain a significant improvement of localization with non-individual HRTF. Effectively, the results of the two groups with only one adaptation session (*G*1 and *B*1) did not show any significant improvement whereas results of the two other groups (*G*3 and *B*3) revealed an improvement of the localization accuracy.

The majority of adaptation improvement was evident by the observed decrease in polar angle error, linked to spectral cues. At the same time, a slight improvement was seen in lateral errors for participants with "good" HRTFs after several sessions (group *G*3). This improvement in interpreting ITD cues is probably due to an adaptation with regards to the imprecision in the ITD model used for non-individual HRTF synthesis. The same improvement was not found for subjects using "bad" HRTFs, indicating possible difficulties in adaptation to multiple inconsistencies in localization cues in such a short time. The adaptation process to polar angle was effective for all HRTF groups with the same level of improvement after three days of training (approximately 23° on the median error). The differences between the two groups were maintained throughout the experiment with the same difference in accuracy for the tests *L*1 and *L*4 (approximately 8°). At the end of three adaptation sessions, performances of subjects with "good" HRTFs achieved the initial performance levels of control subjects with individual HRTFs.

In order to separate the effect of the adaptation task from any learning effect of the protocol, these results should be moderated by the results of the control group with individual HRTFs. The improvement of the control group was approximately 10° for polar error. This can be attributed to the procedural learning of the protocol and rendering system, since it appeared after the first adaptation session and there was almost no change thereafter. The global polar improvement of groups *G*3 and *B*3 was approximately 23° each, implying that the improvement attributed to the adaptation task alone would be 13°. As seen in Hofman *et al.*,[18] there was an improvement in localization accuracy due to a training procedure. In contrast to that study, which presented adaptation to modified pinnae after several days, the adaptation in this study was achieved using a virtual environment and three sessions of 12 min. For both non-individual HRTF groups confusion errors were reduced. A reduction of all confusion types was found for the "bad" HRTF group while the "good" HRTF group principally reduced the number of front/back confusions. This result is comparable to the results of Zahorik *et al.*[27] who obtained a reduction from 38% to 23% in front/back confusions after two training sessions of 30 min with auditory, visual, and proprioceptive/vestibular feedback.

Compared to previous studies, the general results of this study is that naive subjects using individualized HRTFs are slightly poorer than results presented by Wightman *et al.*,[13] who reported 11% front/back confusions vs 14% in the current study. Considering the performance of subjects using non-individualized HRTF, the results of Wenzel *et al.*[3] showed 31% front/back confusion rate (29% of which were combined confusion errors) and 18% up/down confusion rate (55% of which were combined confusion errors). Separating the combined confusion errors, these results equate to an error type distribution of 22% front/back confusion and 8% of up/down confusion, which is quite comparable to the results obtained respectively by groups *G*3 and *B*3 in this study: 22% and 25% for front/back confusions with 7% and 11% for up/down confusions. The exact method used to define the error zones in previous studies differed from the method used in the current study, especially for combined errors, making precise comparisons impossible, but it is clear that the results are highly comparable, which is sufficient for the purposes of this analysis.

V. CONCLUSION

The results from the current study demonstrate that a rapid perceptual adaptation to non-individual HRTF through an audio proprioceptive and vestibular feedback is possible. Moreover, it shows that the adaptation of auditory system does not necessarily need visual feedback and can be properly achieved through other senses in the full auditory sphere of perception.

This study only focused on three sessions of adaptation. Regarding the results, it seems clear that more sessions would lead to additional improvement. It is not certain how many training sessions would be required to completely adapt to the non-individual HRTFs, but it is possible to imagine that after a certain number of adaptation sessions, localization performance with non-individual virtual cues would coincide with a free field listening situation. This would occur faster for "good" match HRTFs, highlighting the benefit of HRTF selection methods (see Katz and Parseihian[31]).

There are many applications where the use of virtual auditory displays is hindered by the problem of front/back reversals, such as with orientation and navigation tasks. These results are significant for developers and users of VAE. Whereas HRTF individualization to a given user requires significant measurements and/or computational processing, the method presented here gradually adapts the user to the

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22

non-individual display using a best selection HRTF as a starting point. On the one hand, the combination of individualized ITD cues (with the measurement of head circumference) with a selected HRTF set (using a perceptual judgment) allows for the creation of an optimized hybrid HRTF that matches the subject. On the other hand, three or more training sessions using a multimodal platform allow the user to reach the performance of individualized HRTF, and additional training can further improve precision. The use of a non-visual adaptation procedure enables this solution to be applied to visually impaired users, allowing better acceptability and performance of emerging assistive technologies based on spatial audio.

ACKNOWLEDGMENTS

This work was supported in part by the French National Research Agency (ANR) through the TecSan program (project NAVIG ANR-08TECS-011) and the Midi-Pyrénées region through the APRRTT program. The authors would like to thank Alan Blum and Bastien Francony for the previous studies on this experiment. Finally, the authors would like to thank the reviewers for their thoughtful and precise comments.

[1]J. Blauert, *Spatial Hearing* (MIT Press, Cambridge, 1996), pp. 36–200.
[2]D. R. Begault, *3-D Sound for Virtual Reality and Multimedia* (Academic Press, Cambridge, 1994), pp. 117–190.
[3]E. M. Wenzel, M. Arruda, D. J. Kistler, and F. L. Wightman, "Localization using nonindividualized head-related transfer functions," J. Acoust. Soc. Am. **94**(1), 111–123 (1993).
[4]J. C. Middlebrooks, "Individual differences in external-ear transfer functions reduced by scaling in frequency," J. Acoust. Soc. Am. **106**(3), 1480–1492 (1999).
[5]J. C. Middlebrooks, "Virtual localization improved by scaling nonindividualized externalear transfer functions in frequency," J. Acoust. Soc. Am. **106**(3), 1493–1510 (1999).
[6]S. L. Lee, L. H. Kim, and K. M. Sung, "Reduction of sound localization error for nonindividualized HRTF by directional weighting function," IEICE Trans. Fund. Electron. Commun. Comput. Sci. **87**(6), 1531–1536 (2004).
[7]Y. Kahana and P. A. Nelson, "Boundary element simulations of the transfer function of human heads and baffled pinnae using accurate geometric models," J. Sound Vib. **300**(3–5), 552–579 (2007).
[8]B. F. G. Katz, "Boundary element method calculation of individual head-related transfer function. II. Impedance effects and comparisons to real measurements," J. Acoust. Soc. Am. **110**(5), 2440–2448 (2001).
[9]Y. Iwaya, "Individualization of head-related transfer functions with tournament-style listening test: Listening with other's ears," Acoust. Sci. Technol. **27**(6), 340–343 (2006).
[10]B. U. Seeber and H. Fasti, "Subjective selection of non-individual head related transfer function," in *Int. Conf. on Auditory Display* (ICAD Boston, MA, 2003), pp. 259–262.
[11]V. R. Algazi, C. Avendano, and R. O. Duda, "Estimation of a spherical-head model from anthropometry," J. Audio Eng. Soc. **49**(6), 472–479 (2001).
[12]V. R. Algazi, C. Avendano, and R. O. Duda, "Elevation localization and head-related transfer function analysis at low frequencies," J. Acoust. Soc. Am. **109**(3), 1110–1122 (2001).
[13]F. L. Wightman and D. J. Kistler, "Headphone simulation of free-field listening. II: Psychophysical validation," J. Acoust. Soc. Am. **85**(2), 868–878 (1989).
[14]F. L. Wightman and 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).
[15]E. I. Knudsen, "The role of auditory experience in the development and maintenance of sound localization," Trends Neurosci. **7**(9), 326–330 (1984).
[16]A. S. Keuroghlian and E. I. Knudsen, "Adaptive auditory plasticity in developing and adult animals," Prog. Neurobiol. **82**(3), 109–121 (2007).
[17]P. T. Young, "Auditory localization with acoustical transposition of the ears," J. Exp. Psychol. **11**(6), 399–429 (1928).
[18]P. M. Hofman, J. G. Van Riswick, and A. J. Van Opstal, "Relearning sound localization with new ears, Nat. Neurosci. **1**(5), 417–421 (1998).
[19]M. P. Zwiers, A. J. Van Opstal, and G. D. Paige, "Plasticity in human sound localization induced by compressed spatial vision," Nat. Neurosci. **6**(2), 175–181 (2003).
[20]M. P. Zwiers, A. J. Van Opstal, and J. R. Cruysberg, "A spatial hearing deficit in early-blind humans," J. Neurosci. **21**(9), 1–5 (2001).
[21]M.-E. Doucet, J.-P. Guillemot, M. Lassonde, J.-P. Gagné, C. Leclerc, and F. Lepore, "Blind subjects process auditory spectral cues more efficiently than sighted individuals," Exp. Brain Res. **160**(2), 194–202 (2005).
[22]N. Lessard, M. Paré, F. Lepore, and M. Lassonde, "Early-blind human subjects localize sound sources better than sighted subjects," Nature **395**(6699), 278–280 (1998).
[23]J. Lewald, "Vertical sound localization in blind humans," Neuropsychologia **40**(12), 1868–1872 (2002).
[24]M. P. Zwiers, A. J. Van Opstal, and J. R. Cruysberg, "Two-dimensional sound-localization behavior of early-blind humans," Exp. Brain Res. **140**(2), 206–222 (2001).
[25]B. G. Shinn-Cunningham, N. I. Durlach, and R. M. Held, "Adapting to supernormal auditory localization cues. I. Bias and resolution," J. Acoust. Soc. Am. **103**(6), 3656–3666 (1998).
[26]B. G. Shinn-Cunningham, N. I. Durlach, and R. M. Held, "Adapting to supernormal auditory localization cues. II. Constraints on adaptation of mean response," J. Acoust. Soc. Am. **103**(6), 3667–3676 (1998).
[27]P. Zahorik, P. Bangayan, V. Sundareswaran, K. Wang, and C. Tam, "Perceptual recalibration in human sound localization: learning to remediate front-back reversals." J. Acoust. Soc. Am. **120**(1), 343–359 (2006).
[28]A. Honda, H. Shibata, J. Gyoba, K. Saitou, Y. Iwaya, and Y. Suzuki, "Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game," Appl. Acoust. **68**(8), 885–896 (2007).
[29]A. Blum, B. F. G. Katz, and O. Warusfel, "Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training," in *Proc. CFA/DAGA* (2004).
[30]IRCAM LISTEN HRTF database, available at http://recherche.ircam.fr/equipes/salles/listen/ (Last viewed 10/28/11).
[31]B. F. G. Katz and G. Parseihian, "Percptually based head-related transfert function database optimization," J. Acoust. Soc. Am. **131**(2), EL99–EL105 (2012).
[32]B. F. G. Katz, E. Rio, and L. Picinali, "LIMSI Spatialization Engine," Inter Deposit Digital Number: F. 001.340014.000.S.P. 2010.000.31235.
[33]L. Haber, R. N. Haber, S. Penningroth, K. Novak, and H. Radgowski, "Comparison of nine methods of indicating the direction to objects: Data from blind adults," Perception **22**(1), 35–47 (1993).
[34]F. Dramas, B. F. G. Katz, and C. Jouffrais, "Auditory-guided reaching movements in the peripersonal frontal space (poster)," in *Acoustics 08, Paris, 29/06/2008-04/07/2008*, J. Acoust. Soc. Am. **123**, 3723 (2008).
[35]M. Morimoto and H. Aokata, "Localization cues of sound sources in the upper hemisphere," J. Acoust. Soc. Jpn. **5**(3), 165–173 (1984).
[36]E. M. Wenzel, F. L. Wightman, and D. J. Kistler, "Localization with non-individualized virtual acoustic display cues," in *CHI'91: Proceedings of the SIGCHI Conference on Human factors in Computing Systems* (ACM, New York, 1991), pp. 351–359.
[37]R. L. Martin, K. I. McAnally, and M. A. Senova, "Free-field equivalent localization of virtual audio," J. Audio Eng. Soc. **49**(1/2), 14–22 (2001).
[38]D. Yamagishi and K. Ozawa, "Effects of timbre on learning to remediate sound localization in the horizontal plane," in *Principles and Applications of Spatial Hearing* (World Scientific, Singapore, 2011).

Redistribution subject to ASA license or copyright; see http://acousticalsociety.org/content/terms. Download to IP: 192.236.36.29 On: Thu, 20 Nov 2014 12:28:22