ARCHIVES OF ACOUSTICS
Vol. **40**, No. 4, pp. 491–502 (2015)

DOI: 10.1515/aoa-2015-0049

Multimodal Perceptual Training for Improving Spatial Auditory Performance in Blind and Sighted Listeners

Oana BĂLAN, Alin MOLDOVEANU, Florica MOLDOVEANU

Faculty of Automatic Control and Computers
University POLITEHNICA of Bucharest
Splaiul Independentei, 313, 060042, Bucharest, Romania; e-mail: oana.balan@cs.pub.ro

(*received May 25, 2015; accepted September 25, 2015*)

The use of individualised Head Related Transfer Functions (HRTF) is a fundamental prerequisite for obtaining an accurate rendering of 3D spatialised sounds in virtual auditory environments. The HRTFs are transfer functions that define the acoustical basis of auditory perception of a sound source in space and are frequently used in virtual auditory displays to simulate free-field listening conditions. However, they depend on the anatomical characteristics of the human body and significantly vary among individuals, so that the use of the same dataset of HRTFs for all the users of a designed system will not offer the same level of auditory performance. This paper presents an alternative approach to the use on non-individualised HRTFs that is based on a procedural learning, training, and adaptation to altered auditory cues. We tested the sound localisation performance of nine sighted and visually impaired people, before and after a series of perceptual (auditory, visual, and haptic) feedback based training sessions. The results demonstrated that our subjects significantly improved their spatial hearing under altered listening conditions (such as the presentation of 3D binaural sounds synthesised from non-individualized HRTFs), the improvement being reflected into a higher localisation accuracy and a lower rate of front-back confusion errors.

Keywords: front-back confusions, HRTF, sound localization, training, virtual auditory environment.

1. Introduction

The human auditory system decodes the position of a sound source in space by using a set of acoustic cues that are enclosed in the Head Related Transfer Function (HRTF), a complex function of frequency which describes how a sound wave is filtered by the head and pinna before entering the inner ear. The sound localication cues contained in the HRTF spectrum are related with the lateralisation of the source and with elevation discrimination (Parseihian, Katz, 2012). As in the literature the term "lateralisation" has been used to describe the identification of virtual sound sources rendered over headphones, in this study we will use the term "localisation" to indicate the discrimination of target sound sources in virtual auditory displays. As pointed by Furman *et al.* (2013), the argument in favor of this change of terminology is that "the sound scene is localised by the listener outside the head", similarly to the sound presentation over loudspeakers.

Spatialised sound is very important for producing high-quality acoustic effects and for increasing the user's sense of presence in virtual auditory environments (Meshram *et al.*, 2014). The binaural 3D sound is obtained by filtering the input wave with the corresponding HRTF of a given position in space, with the purpose of simulating free-field listening conditions in virtual auditory environments presented over stereophonic headphones. As the HRTFs significantly depend on the anatomical characteristics of the external auditory system (the size and shape of the pinna, head, and torso), they noticeably vary among individuals, being highly individualised to each listener apart (Dellepiane *et al.*, 2008). Therefore, using non-individualised HRTFs in virtual auditory environments results in localisation errors, lateralisation artifacts, perceptual distortions, such as front-back and back front-confusions (situation when the listener perceives the sound coming from the front as originating from the back and vice-versa) and non-externalized auditory images (Parseihian, Katz, 2012; Meshram *et al.*, 2014).

As generating individualised sets of HRTFs involves a long and complicated measurement process, in order to minimise the sound localisation errors and help listeners to adapt to virtual auditory environments, many experiments have used training methods aimed at recalibrating the auditory system to altered hearing conditions and improving perceptual learning (Parseihian, Katz, 2012; Mendonça, 2014). The adaptive plasticity that occurs after the loss of a sensory modality (especially sight) is the result of multimodal perceptual cues and spatial attention (the ability to focus on specific stimuli of the environment) (Strelnikov *et al.*, 2011).

Other experiments test the pitch discrimination and pitch-timbre categorisation abilities (Bogusz *et al.*, 2012a), pitch memorisation skills (Bogusz *et al.*, 2012b), or discuss the results of auditory training for blind and visually impaired children and teenagers (Bogusz-Witczak *et al.*, 2015). Moreover, other research studied the effect of 3D sound recording and reproduction methods on performance in sound localisation by normally sighted and visually impaired subjects (Furmann *et al.*, 2013).

In our research, we focus on the theory according to which auditory recalibration can be achieved by using a method based on multimodal interaction that forces the auditory system to adapt to the perception of non-individualised HRTFs (Blum *et al.*, 2004).

This paper presents a comparative study on the degree of spatial auditory improvement achieved by a group of sighted and visually impaired individuals who participated in a sound localisation experiment that used a multimodal perceptual feedback based training procedure. The sound localisation performance was assessed before, after, and during the training session. The results demonstrated that both the sighted and visually impaired subjects succeeded to improve their spatial auditory resolution that has been reflected in a higher angular precision accuracy and a lower rate of front-back confusions.

2. Training sound localisation

Sound localisation does not depend only on the auditory cues. By contrary, it is a complex phenomenon which involves the interaction of multimodal sensory cues (vision, proprioception), a certain degree of plasticity achieved as a result of experience and the presence of dynamic cues (Moldoveanu, Bălan, 2014). Perceptual learning represents the performance obtained from practice or experience that improves a person's capacity to interact with the environment. It has been demonstrated in different experiments that sound localisation accuracy can be enhanced through perceptual learning, training, and behavioral adaptation (Honda *et al.*, 2013; Bălan *et al.*, 2014). Attention is a key factor that triggers plasticity and adaptation of the auditory system. It therefore initiates complex dynamic changes in the functionality of the cortical receptive fields, enables the separation of various stimuli and emphasises the salient acoustic features of the environment. Moreover, attention is responsible for updating the sensory filtering characteristics of the auditory system in respect to the listening conditions and the challenges it has to face. Also, attention generates a plasticity effect in the primary auditory cortex that involves a sudden reshape of the spectro-temporal receptive fields of single neurons in the primary auditory cortex (Fritz *et al.*, 2007; King *et al.*, 2011; Ahissar, 2001).

In the majority of the studied experiments, the auditory recalibration procedure was based on proprioceptive feedback (Honda *et al.*, 2013), audiovisual (Strelnikov *et al.*, 2011; Shinn-Cunningham *et al.*, 2005) and multimodal training (Blum *et al.*, 2004), or 3D audio games (Honda *et al.*, 2007; Blum *et al.*, 2004). Most of the subjects reached the same level of localisation performance as their counterparts from the control group (who listened to 3D sounds filtered with individualised HRTFs) after performing several training sessions (Parseihian, Katz, 2012; Blum *et al.*, 2004). Moreover, the training procedure proved to have long-term effects, as the listeners recorded similar levels of performance even one month after the experiment had taken place (Honda *et al.*, 2007). In addition, the game-based training method helped the subjects to adapt to altered hearing conditions (the perception of 3D binaural sounds synthesised with non-individualised HRTFs in virtual auditory environments) and to recalibrate their spatial auditory representation while focusing on the most salient features of the incoming sound (Ohuchi *et al.*, 2006).

Many sound localisation experiments demonstrated that visually impaired individuals can perform sound localization tasks with equal or better accuracy than sighted people. Thus, Lessard *et al.* (1998) suggested that in the case of the visually impaired people compensation takes place through the remaining senses, in particular the hearing sense. Moreover, half of the blind subjects who participated in the study of Doucet (Doucet *et al.*, 2005) showed enhanced auditory discrimination abilities under monaural listening conditions, having one ear blocked. Also, the visually impaired individuals proved to have an enhanced capacity (considered by the authors as supra-normal) of discriminating and processing the spectral content of the sound. There are also other studies which demonstrated that the blind people are able to create a solid spatial representation of the environment and to perform with high accuracy in specific tasks that involve the discrimination of slight changes in the spectral characteristics of the sound (Ohuchi *et al.*, 2006; Röder *et al.*, 1999; Wersényi, 2012). In Zwiers's experiment (Zwiers *et al.*, 2001) the azimuth localisation accuracy of the blind subjects (when they were exposed to long-duration, broadband Gaussian white noise stimuli) was comparable with that of the sighted individuals. Moreover, the elevation localisation accuracy was similar between the two groups of subjects, independent of the pointing method that has been employed. In the experiment designed by Katz and Picinali (2011), there was no significant difference between the blind and the normally sighted group in what concerns the absolute distance error (the difference between the correct and the selected location of the sound source). However, the congenitally blind people recorded a lower angular precision accuracy than the late-blind and the sighted subjects.

3. Method

3.1. Overview

This study comprised several sound localisation tasks. For both groups of subjects (sighted and visually impaired), the experimental procedure included a pre-test session (where the spatial auditory resolution of the subjects has been evaluated in a virtual auditory environment based on 3D binaural sounds synthesised with non-individualised HRTFs), a training session (aimed at helping the subjects to adapt to altered listening conditions) and a post-test session, identical with the pre-test procedure, with the purpose of assessing the level of improvement achieved as a result of training (Fig. 1).

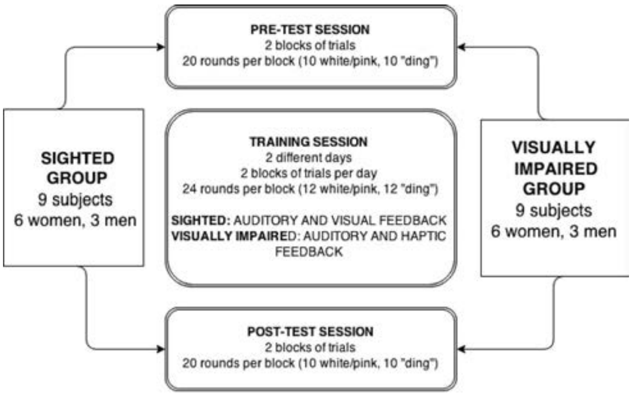


Fig. 1. Experimental procedure.

3.2. Participants

Nine sighted subjects (6 women and 3 men, aged 13–26, mean age 17.5 years) and another nine visually impaired individuals (6 women and 3 men, aged 27–52, mean age 42 years, with a percent of residual vision ranging from 0% to 20%. One subject was congenitally blind (0% residual vision), another was congenitally visually impaired (15% residual vision), while the other 7 were late-onset visually impaired, suffering from certain forms of visual impairments for 6 to 20 years) took part in our experiment. Both groups reported normal hearing and the sighted subjects reported normal or corrected-to-normal vision.

3.3. Ethics statement

All the participants were informed about the aim of the experiment and gave their written consent before the start of the tests. Both experiments were conducted in accordance with the principles stated in the 1964 Declaration of Helsinki and the resulting data were processed anonymously.

3.4. Sound stimuli

For both the training and the test sessions, the sound stimuli consisted of continuous, computer-generated 3D binaural sounds that have been synthesised using the HRTF pairs corresponding to the simulated source position in space. The HRTF set used was taken from the MIT database (MIT HRTF database). The first sound stimuli were a train of broadband white and pink noise that were perceived simultaneously, but in different proportions, according to the direction of the sound source in space. Thus, the listener perceived white noise for the sources situated in the front, pink noise for the sources located in the back, and gradient-varying levels of white and pink noise for the lateral positions. This spectral coloration method is aimed at helping the listeners to differentiate between the sounds situated in the front and in the rear and has as main purpose reducing the incidence of reversal errors (front-back and back-front confusions). The formula for calculating the proportion of white and pink noise corresponding to a certain angular direction in space is the following:

$$\begin{gathered} pink = angle/180; \quad white = 1 - pink; \\ (0 \leq angle \leq 180) \\ white = (angle - 180)/180; \quad pink = 1 - white; \\ (180 < angle < 360). \end{gathered} \tag{1}$$

The second sound stimulus was a repetitive "ding" type soundscape (250 ms pause between each signal) with a narrower spectral profile (Csapó, Wersényi, 2013, Fig. 2). The auditory stimuli were delivered

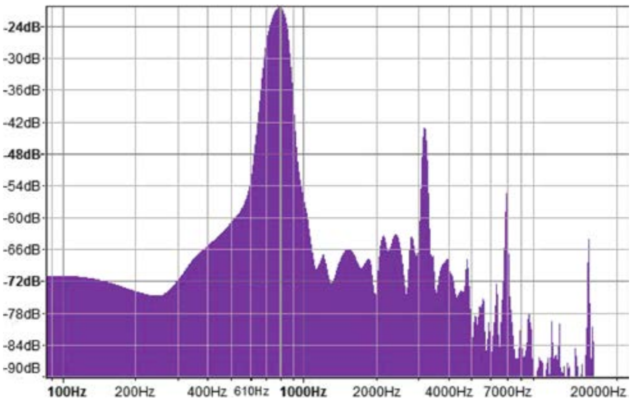


Fig. 2. Spectral profile of the "ding" sound with a peak at 1000 Hz.

through a pair of stereophonic headphones. The headphones used were the Sony MDRZX310L open headphones with no external correction of frequency characteristics. The presented level of the sounds was set to be confortable for the listener, having on average around 65–70 dB SPL.

3.5. Procedure

In the pre-test session, the sound localisation performance of the subjects was tested in a virtual auditory environment where the listeners were required to identify the location of a hidden target, based on the perceived auditory cues. There have been two blocks of trials, each block containing 20 rounds (10 rounds using the white and pink noise and another 10 using the "ding" sound). The listeners were asked to navigate freely, having the mouse movements as the main interaction modality (although 3 visually impaired people used the touchpad), from the starting position to the location of the target sound source in a 2D setting. As the users were modifying their virtual position in respect with the target source, the auditory stimuli that they perceived through the headphones were changing in both the spectral content (corresponding to the direction of the sound source) and amplitude. Thus, as the listeners got nearer to target, the intensity of the sound increased, while, as they got farther, the perceived volume decreased until total silence (outside the auditory range of 200 pixels). The formula that calculates the perceived volume of the target sound source is the following:

$$
\mathit{Gain\ factor} = \begin{cases} 0, & d > dmax \\ GFMIN + (GFMAX - GFMIN) & \\ \quad * \left(1 - \dfrac{d}{dmax}\right)^2, & d \le dmax \end{cases} \tag{2}
$$

where d is the current distance between the position of the listener and the target sound source, $dmax = 150$ pixels, $GFMIN = 0.05$ (the minimum gain factor), $GFMAX = 1$ (the maximum gain factor).

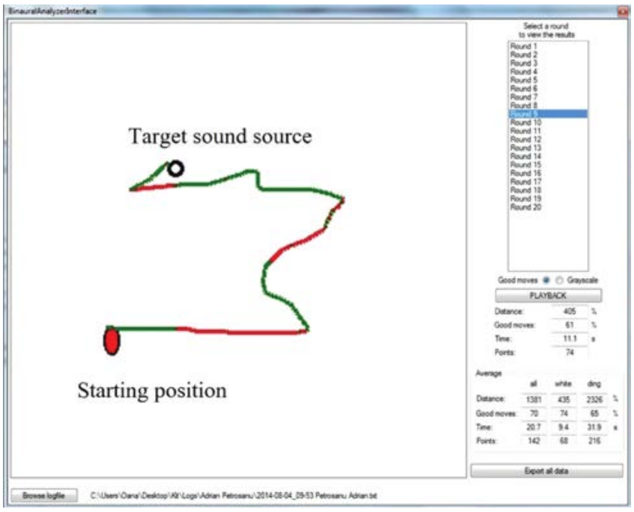
At each round, when the subjects reached the target (defined as a single point location), they could hear a symbolic auditory cue (a sparkling sound) that informed them that the task has been currently completed.

The studied parameters were:

P1: The ratio of distance travelled by the listener to the minimum possible distance (Fig. 3).

P2: The percent of correct travel decisions, defined as movements effectuated towards the sound source, minimising the distance between the user's virtual location and the position of the target (Fig. 3).

P3: The round completion time (in seconds), i.e. the amount of time that passes from the moment the listener starts to hear the sound until he/she finally reaches the target.



Good moves ● ○ Grayscale

Distance: 405 %
Good moves: 61 %
Time: 11.1 s
Points: 74

Average

	all	white	ding	
Distance:	1381	435	2326	%
Good moves:	70	74	65	%
Time:	20.7	9.4	31.9	s
Points:	142	68	216	

C:\Users\Oana\Desktop\All\Logs\Adrian Petrosanu\2014-08-04_09-53 Petrosanu Adrian.txt

Fig. 3. Path travelled by the user from the starting point to the target sound source. The correct travel decisions are coloured in green and the wrong ones are red.

The training session consisted of 4 blocks of trials (each block having 24 trials, 12 based on white and pink noise, and 12 using the "ding" sound), separated into 2 different days (2 blocks of trials in each day). In these trials, both groups of subjects were required to listen to continuous 3D binaural sounds delivered through headphones and to indicate the perceived location of the sound source in space.

The sighted subjects were offered auditory and visual feedback, as the correct direction of the sound was drawn on the screen (coloured in green), together with the listener's choice (coloured in red), and a simulation of the target sound was rendered over headphones in order to recalibrate the spatial perception and create a solid multimodal association between the visual and the hearing senses. On the other hand, the visually impaired people were offered auditory and haptic feedback. They were required to listen to various sound stimuli (corresponding to the directions 0, 30, 60, 90, 120, 150, 180, 210, 240, 270, 300, and 330 degrees) and indicate the perceived direction using the convention of the hour hand of the clock (for instance, 0 degrees to the front corresponded to the 12 o'clock, while 90 degrees to the right was associated to the 3 o'clock). After the visually impaired subjects indicated their response to the examiner, they consequently received auditory (the target sound was emulated over headphones again) and haptic feedback, through a series of positional vibrations (corresponding to the direction of the target sound source) that the listener perceived on the haptic belt he/she was required to wear on the head. The haptic belt consisted of 12 vibration motors placed at 30 degrees distance all around the listener's head (Fig. 4).

Fig. 4. Visually impaired subject during the training session.

For the training session, the studied parameters were the reversal error rate (the percent of front-back and back-front confusions for each block of trials) and the mean angular precision error (mean unsigned error, defined as the difference between the correct direction of the sound source and the direction perceived by the listener).

The post-test session was carried out using exactly the same stimuli and sound localisation procedures as in the pre-test session. All experiments took place in a quiet room, without any interference from outside.

4. Results

4.1. Pre-test and the post-test sessions

For the rounds where the combination of white and pink noise in varying proportions according to the direction of the sound source in space has been used, the mean value of parameter P1 (ratio of the distance travelled by the listener to the minimum possible distance between the starting position and the target location) improved with 24.6% for the sighted group (although the results are not significant at $p < 0.1$) and with 32% for the visually impaired group (in a Wilcoxon Signed Rank Test at $p < 0.05$, Table 1). The results of the sighted subjects indicated a better evolution in the mean values of P1 (as compared to their visually impaired counterparts), for both the pre-test and the post-test sessions of the experiment. (We performed an omnibus analysis followed by a series of post-hoc tests. Thus, for the pre-test session, the results are significant in an ANOVA test, $p = 0.014$ and in a Student t-test where $t = 2.74$, $p = 0.007$. For the post-test session, the results are significant in the ANOVA test, $p = 0.03$ and in a Student t-test where $t = 2.37$ at $p = 0.015$). Moreover, all the nine visually impaired subjects succeeded to obtain improvements in the mean value of P1 in the post-test session, whereas only 66% of the sighted participants recorded a higher level of performance for this parameter (Fig. 5).

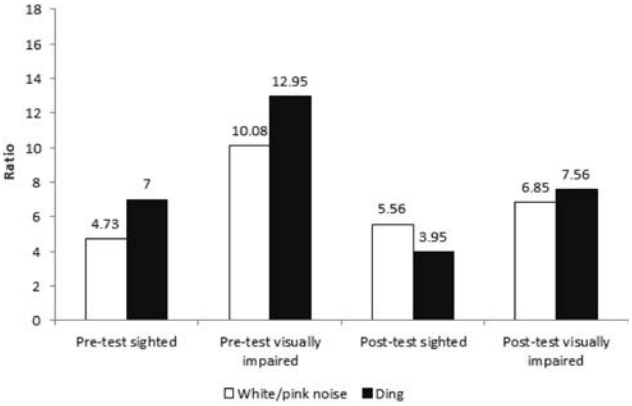


Fig. 5. Evolution of parameter P1 in the pre-test and post-test sessions, for the sighted and the visually impaired groups.

Table 1. Results of the pre-test and post-test sessions.

Type of sound	Parameter	Sighted				Visually Impaired			
		Pre-test	Post-test	Level of improvement between the pre-test and the post-test session [%]	Percent of subjects who obtained an improvement in the post-test session (%)	Pre-test	Post-test	Level of improvement between the pre-test and the post-test session [%]	Percent of subjects who obtained an improvement in the post-test session [%]
White/ pink	P1	4.7	3.5	24.6	66	10	6.8	32	100
	P2 (%)	72.4	77.3	6.7	66	64	66.1	3.2	77
	P3 (seconds)	19.2	17.4	9.2	88	26.6	22	17.1	88
Ding	P1	7	3.9	43.6	100	12.95	7.5	41.6	100
	P2 (%)	70.6	74.6	5.5	77	61.2	65.1	6.3	100
	P3 (seconds)	24.8	16.9	31.8	100	32.8	23.5	28.2	88

In what concerns parameter P2 (the percent of correct travel decisions towards reducing the distance to the sound source), the mean value increased with 6.7% (from 72.4% to 77.3%) for the sighted group (the results are significant in an ANOVA test, $p = 0.08$ and in a Student t-test for dependent means where $t = 1.93$ at $p = 0.08$) and with 3.2% (from 64% to 66.1%) in the case of the visually impaired subjects (the results are significant in a Student t-test where $t = 2.46$ at $p = 0.038$) (Fig. 6). The sighted individuals outperformed the visually impaired subjects in both the pre-test and the post-test sessions of the experiment. (For the pre-test session, the results are significant in an ANOVA test, $p = 0.002$ and in a Student t-test where $t = 3.55$, $p = 0.0013$. For the post-test session, the results are significant in the ANOVA test at $p = 8.9\text{E-}5$ and in a Student t-test, where $t = 5.18$ at $p = 4.5\text{E-}5$). Similarly to P1, a higher percent of the visually impaired participants were able to improve their correct travel decision rate between the pre-test and the post-test sessions (77%, compared to 66% for the sighted group).

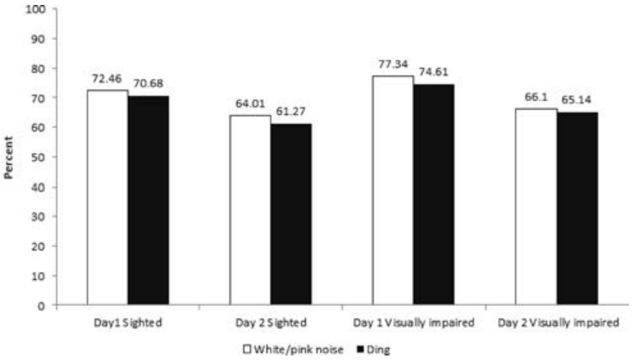


Fig. 6. Evolution of parameter P2 in the pre-test and post-test sessions, for the sighted and the visually impaired groups.

Regarding parameter P3 (the round completion time), its mean value decreased with 9.2% in the post-test phase (from 19.2 seconds to 17.4 seconds) in the case of the sighted subjects (although the results are not statistically significant at $p < 0.1$) and with 17.1% (from 26.6 seconds to 22 seconds) for the visually impaired group (the results are not statistically significant at $p < 0.1$). The sighted subjects outperformed their visually impaired counterparts in the pre-test session ($t = 1.35$, $p = 0.09$), although the results of both groups converged in the post-test stage (in an ANOVA test, the differences were not statistically significant at $p < 0.1$). However, in both groups, 8 of 9 subjects (88%) recorded a noticeable improvement as a result of perceptual adaptation for parameter P3.

For the rounds that used the narrowband "ding" signal as the main auditory cue, the mean value of parameter P1 improved with 43.6% for the sighted group (the results are statistically significant in an ANOVA, $p = 0.05$ and in a Student t-test where $t = -2.83$ at $p = 0.02$) and with 41.6% for the visually impaired group (the results are statistically significant in an ANOVA test, $p = 0.04$ and in a Student t-test where $t = -4.44$, $p = 0.001$). The results recorded by the sighted subjects showed a better evolution in what concerns the mean values of P1 (compared to their visually impaired individuals), for both the pre-test and the post-test sessions (for the pre-test session, the results are significant in an ANOVA test, $p = 0.03$ and in a Student t-test where $t = 2.29$, $p = 0.017$. For the post-test session, the results are significant in the ANOVA test, $p = 0.01$ and in a Student t-test where $t2.7$ at $p = 0.007$). At the same time, all the sighted and the visually impaired subjects (100%) succeeded to reduce the ratio of the distance travelled to the minimum possible distance between the starting position and the target sound source during the post-test trials.

Regarding parameter P2, the mean value increased by 5.5% (from 70.6% to 74.6%) for the sighted group (the results are significant in an ANOVA test, $p = 0.09$ and in a Student t-test where $t = 2.71$ at $p = 0.026$) and by 6.3% (from 61.2% to 65.1%) in the case of the visually impaired subjects (the results are statistically significant in a Wilcoxon Signed Rank Test at $p < 0.05$). The sighted individuals' percent of correct travel decisions was higher than that of the visually impaired subjects in both the pre-test and the post-test sessions (for the pre-test session, the results are significant in an ANOVA test, $p = 0.0001$ and in a Student t-test where $t = 4.4$, $p = 0.00002$. For the post-test session, the results are significant in the ANOVA test, $p = 0.002$ and in a Student t-test where $t = 3.67$ at $p = 0.001$). Similarly to the rounds where the white/pink noise combination has been used, a higher percent of the visually impaired subjects improved the mean rate of parameter P2 after the training procedure (100% for the visually impaired group and 77% for the sighted group).

In what concerns parameter P3, the decrease in the round completion time was 31.8% (from 24.8 seconds to 16.9 seconds) for the sighted group (the results are statistically significant in an ANOVA test, $p = 0.02$ and in a Student t-test for dependent means where $t = -3.37$ at $p = 0.004$) and 28.2% (from 32.8 seconds to 23.5 seconds) for the visually impaired group ($t = 2.1$ at $p = 0.03$). In this case, the sighted subjects achieved better results for the mean round completion time than the visually impaired subjects in the pre-test session, although the results of both groups were not statistically different. Nonetheless, all the sighted subjects succeeded to complete the rounds quicker in the post-test trials, while only 8 of the 9 visually impaired users (88%) obtained a lower mean value of parameter P3 between the two test sessions (Table 2).

Table 2. Statistical comparison between the sighted and the visually impaired subjects in the pre-test and post-test sessions of the experiment.

	Parameter P1		Parameter P2		Parameter P3	
Comparison between the sighted and the visually impaired groups	White-pink noise	"Ding" sound	White-pink noise	"Ding" sound	White-pink noise	"Ding" sound
Pre-test session	ANOVA $p = 0.014$ Student t-test $t = 2.74$ $p = 0.007$	ANOVA $p = 0.03$ Student t-test $t = 2.29$ $p = 0.017$	ANOVA $p = 0.002$ Student t-test $t = 3.55$ $p = 0.0013$	ANOVA $p = 0.0001$ Student t-test $t = 4.4$ $p = 0.0002$	Student t-test $t = 1.35$ $p = 0.09$	No significant difference between the groups
Post-test session	ANOVA $p = 0.03$ Student t-test $t = 2.37$ $p = 0.015$	ANOVA $p = 0.01$ Student t-test $t = 2.7$ $p = 0.007$	ANOVA $p = 8.9\text{E-}5$ Student t-test $t = 5.18$ $p = 4.5E - 5$	ANOVA $p = 0.002$ Student t-test $t = 3.67$ $p = 0.001$	No significant difference between the groups	No significant difference between the groups

4.2. The training session

For the rounds where the combination of white and pink noise in varying proportions according to the direction of sound has been employed, we recorded a reduction of 14.7% in the sound localisation error (from 24.1 to 20.5 degrees) in the case of the sighted subjects and of 25.7% (from 37.3 to 27.7 degrees, Fig. 7) in the case of the visually impaired individuals ($t = 2.57$ at $p = 0.01$, Table 3). Even though the localisation error was larger for the visually impaired group, their improvement rate significantly surpassed that of the sighted counterparts with more than 10%. The same situation is encountered when assessing the front-back

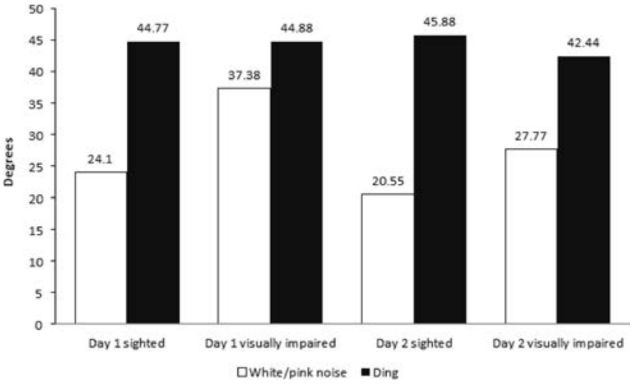


Fig. 7. Evolution of angular precision errors in both days of training, for the sighted and the visually impaired group.

Table 3. Results of the training session.

Type of sound	Parameter	Sighted				Visually Impaired			
		Day 1	Day 2	Level of improvement between first and the second day of training [%]	Percent of subjects who obtained an improvement in the second day of training [%]	Day 1	Day 2	Level of improvement between first and the second day of training [%]	Percent of subjects who obtained an improvement in the second day of training [%]
White/ pink	Angular precision error (degrees)	24.1	20.5	14.7	66	37.3	27.7	25.7	77
	Reversal error rate [%]	8.3	5.5	33.3	77	12	6	50	100
Ding	Angular precision error (degrees)	44.7	45.8	−2.4	44	44.8	42.4	5.4	55
	Reversal error rate [%]	22.7	19.4	14.6	77	14.3	12.4	12.9	55

confusion rate, as the sighted individuals obtained an improvement rate of 33.3% (from 8.3% to 5.5%), while the visually impaired subjects reduced the percent of reversal errors with 50% (from 12% to 6% – the results are statistically significant in an ANOVA test, $p = 0.02$ and in a Student t-test where $t = -3.71$ at $p = 0.002$, Fig. 8). We can also notice that the number of individuals who achieved a higher sound localisation accuracy is larger in the visually impaired group (with 11% more in the case of precision localisation errors and 33% in the case of reversal misjudgments). The visually impaired subjects reported higher angular precision errors than the sighted participants in both days of training (for the first day of training, the differences are statistically significant in an ANOVA test, $p = 0.01$ and in a Student t-test where $t = 2.69$ at $p = 0.008$). On the other hand, the ANOVA test revealed that the front-back localisation judgments are comparable for the sighted and the visually impaired groups, for both days of training (the differences are not significant at $p < 0.1$).

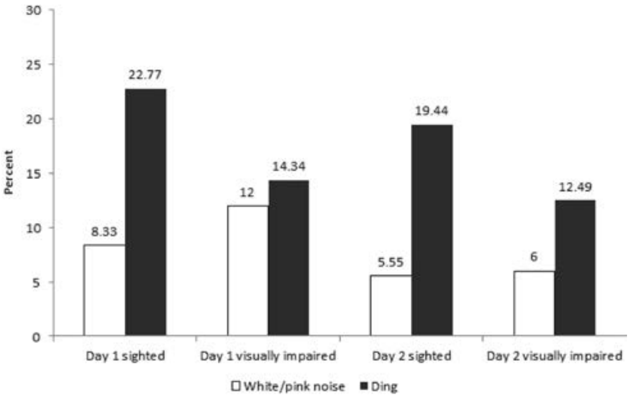


Fig. 8. Evolution of reversal error rates in both days of training, for the sighted and the visually impaired group.

For the rounds that used the “ding” sound as the main auditory stimuli, the sighted subjects recorded an increase in the angular precision error of 2.4% (from 44.7 degrees to 45.8 degrees – the results are not statistically significant), while the visually impaired succeeded to obtain a decrease of 5.4% (from 44.8 degrees to 42.4 degrees). At the same time, the percent of subjects who achieved a better angular localisation accuracy is slightly higher for the visually impaired group (55%, as compared to 44% in the case of the sighted group). In what concerns the front-back confusion rate, the level of improvement is higher for the sighted subjects (14.9%, from 22.7% to 19.4%). Also, the percent of individuals who were able to enhance their front-back auditory localization judgment was larger in the case of the sighted group (77%, as compared to 55% for the visually impaired). The visually impaired subjects achieved a lower rate of front-back confusion errors than the sighted participants in both days of training. (In the first day, the results were significant in an ANOVA test, $p = 0.09$ and in a Student t-test where $t = 1.78$, $p = 0.04$, while in the second day of training the differences between the groups were not statistically significant). Nonetheless, the angular precision performance is comparable for the sighted and the visually impaired groups, in both days of the training session (the differences are not significant at $p < 0.1$, Table 4).

Table 4. Visually impaired subject during the training session.

	Angular precision error		Front-back confusion rate	
Comparison between the sighted and the visually impaired groups	White-pink noise	“Ding” sound	White-pink noise	“Ding” sound
Day 1 of training	ANOVA $p = 0.01$ Student t-test $t = 2.69$ $p = 0.08$ Better performance in the case of the sighted subjects	No significant difference between the groups	No significant difference between the groups	ANOVA $p = 0.09$ Student t-test $t = 1.78$ $p = 0.04$ Better performance in the case of the visually impaired subjects
Day 2 of training	No significant difference between the groups	No significant difference between the groups	No significant difference between the groups	No significant difference between the groups

5. Discussion

The results provided by our experiment demonstrated a rapid improvement in the sound localisation ability and front-back disambiguation for both the sighted and the visually impaired subjects. We consider that the high level of sound localisation improvement and front-back disambiguation is due to learning the procedure how to identify the spectral characteristics of the sound (especially for the white-pink noise combination) and to the perceptual feedback based training method, even if the subjects were still listening to ambiguous stimuli (due to the use of non-individualised HRTFs).

Thus, the sighted subjects created a strong association between the visual and auditory stimuli during the training session, as they could see on the screen the correct direction of the sound source and listen to it at the same time. The visually impaired participants benefited from haptic feedback – they could feel a series of vibrations on the haptic belt they were required to wear on the head (originating from the same direction as the sound source) and simultaneously listen to the 3D target sound delivered over headphones.

In the post-test session of the experiment, the level of improvement recorded by the visually impaired people is higher than that of the sighted subjects (for parameters P1 and P3), for the rounds where the combination of white and pink noise in varying proportions has been used as the main auditory cue. For the rounds where the narrowband "ding" signal has been used, the degree of improvement is higher for the sighted subjects (for parameters P1 and P3), although the visually impaired participants succeeded to obtain a higher percent of improvement for P2 (the rate of correct travel decisions). For both types of sound stimuli, for all the three studied parameters (except for the "ding" sound, for parameter P3), the percent of visually impaired subjects who recorded a significant improvement in the post-test session of the experiment is equal or higher than that of the sighted individuals, demonstrating that the level of spatial auditory adaptation is larger in the case of the subjects who suffer from a certain degree of visual disability.

In the second day of training, the visually impaired outperformed their sighted counterparts in the percent of angular precision and reversal error rate improvement. Also, the number of subjects who obtained a better sound localisation performance as a result of training was higher in the visually impaired group (except for the rounds where the "ding" sound has been used, where the sighted subjects recorded a higher level of improvement than the visually impaired for the reversal error rate parameter).

We believe that the adaptation process that took place during the training session was concentrated on learning how to focus on the spectral characteristics of the auditory stimuli, as an indicator of the spatial position of the target source. Our subjects succeeded to improve their spatial auditory resolution on a rapid time scale, the fact that can be explained by an increased selective attention to the spectral profile of the sounds during the training phase. These results can be explained by the fact that the white and pink noises are more externalised than the narrowband "ding" sound, as the broadband noises contain much more spectral information that facilitates the localisation process. Although they are not very natural sounds, the white and pink noises surpass the narrowband signals due to their enhanced directionality and laterality (externalisation or out-of-the head perception). Also, the white and pink noises were continuous, offering a complete auditory perception, while the "ding" stimuli were discrete and repetitive.

The sighted subjects obtained a higher angular precision accuracy than the visually impaired individuals for the rounds where the white and pink noises have been used as main auditory cues. However, both groups recorded similar reversal error rates, in both days of the training session. In the case of the rounds where the narrowband "ding" signal has been employed, the angular precision error rates were comparable for both groups, although the visually impaired subjects made fewer reversal errors than the sighted participants.

For this type of stimuli, both the sighted and the visually impaired subjects who participated in our experiment obtained comparable reversal error rates, reaching a mean reversal error rate of 5.55% (the sighted), respectively 6% (the visually impaired) in the second day of training. These results demonstrate the efficiency of our method, proving that the combined spectral features of both types of noise enabled the listeners to differentiate between the sources located in the frontal or rear positions.

For the trials where the sonification technique was based on the use of the narrowband "ding" type stimulus, the sound localisation performance of both groups of subjects was weaker than for the rounds which employed the white/pink noise combination. The angular precision error rate was comparable for both groups of subjects. Nonetheless, the visually impaired listeners were able to better disambiguate the sources located on the cone of confusion, reaching a mean reversal error rate of 12.9% in the second day of training. This result demonstrates that the visually impaired individuals are able to identify the hemifield the sound source is originating from with a higher accuracy than the sighted people, even when listening to narrowband auditory stimuli.

Nonetheless, a plausible explanation for the slightly poorer spatial auditory localization accuracy of the visually impaired subjects in the training and post-test session (as compared with that of the sighted individuals) is the limited resolution of the haptic belt used for training (the vibration motors were placed at 30 degrees difference around the head), while the visual feedback (with perfect resolution) used in the case of the sighted individuals provided a complete spatial perception of the environment.

The combination of white and pink noise in varying proportions, according to the direction of the sound source in space led to a higher precision accuracy in the case of the sighted group of subjects. The results of our experiment are better than those obtained by Blum *et al.* (2004), who recorded for their test group a mean precision error of 29 degrees and a rate of front-back confusions of 25% after the training session. Similarly, our results are comparable with those presented by Majdak *et al.* (2010), who recorded a precision error of 23.3 degrees before training and of 19.8 degrees after the visual-auditory feedback based adaptation procedure. In what concerns the front-back confusion rate, our results are better than those obtained by Parseihian and Katz (2012), who recorded a reduction in the front-back confusion rate from 25–27% to 11% in the post-test session of the experiment. The 12 minutes adaptation task (which has been performed for three days) consisted in a game where the subjects were required to search for animal sounds hidden around them, using a hand-held position-tracked ball. Also, we recorded a lower incidence of reversal errors for both groups of subjects than Zahorik *et al.* (2006), who obtained a reduction from 38% to 23% in the front-back confusion rate after 2 training sessions of 30 minutes in which the listeners were provided auditory, visual, and proprioceptive feedback. Moreover, our results are better than those of Wenzel (2001) who recorded a front-back confusion mean rate of 32% (ranging from 20% to 43%) in the virtual auditory environment and comparable with her results under free-field listening conditions (mean reversal error rate of 6.5%, ranging from 2% to 10%). In addition, the performance of our subjects is higher than that of the subjects who participated in Padersen and Jorgensen's experiment (Padersen, Jorgensen, 2005) who recorded a front-back confusion rate of 21.3% for 250 ms long virtual white noise stimuli. Also, our results are even better than the reversal error rate obtained by them under free-field listening conditions (9.1%) for the same type of auditory cues.

6. Conclusions

Our study demonstrated that the human auditory system is able to quickly adapt to altered hearing conditions, such as listening to 3D binaural sounds filtered with non-individualised HRTFs. Both the sighted and the visually impaired listeners succeeded in improving their sound localization performance by reducing the reversal and precision error rates for both types of stimuli (with significant better results for the rounds that employed the synthesis of white and pink noise in varying proportions, according to the direction of the sound source in space). Although the sighted subjects outperformed their visually impaired counterparts in most of the required tasks (the visual resolution is highly more accurate than the spatial resolution of the haptic belt, where the vibration motors have been placed at 30 degrees difference all around the head), the level of improvement of the visually impaired participants in the second day of training, respectively in the post-test session of the experiment was generally higher than that of the sighted individuals. Also, the percent of subjects who recorded a significant improvement as a result of training is higher in the visually impaired group.

In our research, we found evidence that the human auditory system is able to adapt to altered hearing conditions after a short training session based on multimodal (auditory, visual, and haptic) feedback. The results of our test and training sessions prove that the sighted and visually impaired subjects have been able to use 3D binaural sounds synthesised from non-individualised HRTFs as the only means for navigating in a virtual auditory environment. Besides, the results of the post-test session demonstrate that the subjects enhanced their orientation and mobility skills, improved their directional decision-making abilities and recalibrated the spatial resolution when navigating in a virtual auditory environment, for both types of sounds.

In conclusion, the proposed approach can be considered a useful training and rehabilitation tool for the future development of audio-only games and for the design of an assistive device for the blind people.

The following experiments will continue refining this one and providing a more extensive training session aimed at investigating the highest degree of sound localisation accuracy that can be achieved as a result of multimodal perceptual feedback based training. In our future research, we will use a higher resolution haptic feedback. Furthermore, the results that will be obtained will be of significant use for the future development of the training modalities for a Sensory Substitution Device aimed at providing a rich representation of the environment.

Acknowledgment

The work has been funded by the Sectoral Operational Programme Human Resources Development 2007–2013 of the Ministry of European Funds through the Financial Agreement POSDRU/159/1.5/S/132395. This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 643636 "Sound of Vision".

References

1. Ahissar M. (2001), *Perceptual training: A tool for both modifying the brain and exploring it*, PNAS, 98(21).
2. Bălan O., Moldoveanu A., Moldoveanu F., Morar A. (2014), *Experiments on training the human localization abilities*, Proceedings of the 10th International Scientific Conference eLearning and Software for Education-Bucharest, April 24–25, 2014.
3. Blum A., Katz B.F.G., Warusfel O. (2004), *Eliciting adaptation to non-individual HRTF spectral cues with multi-modal training*, Proceedings of Joint Meeting of the German and the French Acoustical Societies (CFA/DAGA '04), Strasboug, France.
4. Bogusz E., Koprowska H., Skrodzka E. (2012), *Investigation of Performance in Selected Psychoacoustic Tasks by Visually Impaired Children and Teenagers*, Acta Physica Polonica A, **121**, 1A, A19–A23.
5. Bogusz E., Koprowska H., Skrodzka E. (2012), *Performance in pitch memory task by visually handicapped children and youths*, Archives of Acoustics, **37**, 4, 549–553.
6. Bogusz-Witczak E., Skrodzka E., Furmann A., Hojan E., Przybek K. (2015), *Results of Auditory Training for Blind and Visually Handicapped Children and Adolescents*, Acta Physica Polonica A, **127**, 117–119.
7. Csapó A., Wersényi Gy. (2013), *Overview of auditory representations in human-machine interfaces*, Journal ACM Computing Surveys (CSUR), **46**, 2.
8. Dellepiane M., Pietroni N., Tsingos N., Asselot M., Scopigno R. (2008), *Reconstructing head models from photographs for individualized 3D-audio processing*, Comput. Graph. Forum, **27**, 7, 1719–1727.
9. Doucet M.E., Guillemot J.P., Lassonde M., Gagné J.P., Leclerc C., Lepore F. (2005), *Blind subjects process auditory spectral cues more efficiently than sighted individuals*, Exp Brain Res. 2005 Jan, **160**, 2, 194–202.
10. Fritz J.B., Elhilali M., David S.V., Shamma S.A. (2007), *Does attention play a role in dynamic receptive field adaptation to changing acoustic salience in A1?*, Hear Res., **229**, 1–2, 186–203.
11. Furmann A., Skrodzka E., Giżewski P., Nowotny Ł. (2013). *Effect of sound reproduction method on performance in sound source localization tasks by visually impaired and normal sighted subjects*, Acta Physica Polonica A, **123**, 6, 988–994.
12. Honda A., Shibata H., Gyoba J., Saitou K., Iwaya Y., Suzuki Y. (2007), *Transfer effects on sound localization performances from playing a virtual three-dimensional auditory game*, Appl. Acoustics, **68**, 8, 885–896.
13. Honda A., Shibata H., Hidaka S., Gyoba J., Iwaya Y., Suzuki Y. (2013), *Effects of head movement and proprioceptive feedback in training of sound localization*, Perception, **4**, 4, 253–264.
14. Katz F.G.B., Picinali L. (2011), *Spatial Audio Applied to Research with the Blind*, Advances in Sound Localization, Dr. Pawel Strumillo [Ed.], ISBN: 978-953-307-224-1, InTech, Available at: http://www.intechopen.com/books/advances-in-sound-localization/spatial-audio-applied-to-research-with-the blind.
15. King A.J., Dahmen J.C., Keating P., Leach N.D., Nodal F.R., Bajo V.M. (2011), *Neural circuits underlying adaptation and learning in the perception of auditory space*, Neurosci Biobehav Rev. 2011 Nov, **35**, 10, 2129–39.
16. Lessard N., Paré M., Lepore F., Lassonde M. (1998), *Early-blind human subjects localize sound sources better than sighted subjects*, Nature, **395**, 278–280.
17. Majdak P., Goupell M.J., Laback B. (2010), *3-D Localization of Virtual Sound Sources: Effects of Visual Environment, Pointing Method, and Training*, Atten Percept Psychophys, **72**, 2, 454–469.
18. Mendonça C. (2014), *A review on auditory space adaptations to altered head-related cues*, Front Neurosci., **2019**, 8, doi: 10.3389/fnins.2014.00219.
19. Meshram A., Mehra R., Yang H., Dunn E., Frahm J.M., Manocha D. (2014), *P-HRTF: Efficient personalized HRTF computation for high-fidelity spatial sound*, International Symposium on Mixed and Augmented Reality.
20. Moldoveanu A., Bălan O. (2014), *Training System for Improving Spatial Sound Localization*, Proceedings of the 10th International Scientific Conference eLearning and software for Education-Bucharest, April 24–25, 2014.
21. Ohuchi M., Iwaya Y., Suzuki Y., Munekata T. (2006), *A comparative study of sound localization acuity of congenital blind and sighted people*, Acoustical Science and Technology, **27**, 290–293.
22. Padersen J.A., Jorgensen T. (2005), *Localization Performance of Real and Virtual Sound Sources*, Proceedings North Atlantic Treaty Organization.
23. Parseihian G., Katz B.F.G. (2012), *Rapid Head-Related Transfer Function adaptation using a virtual auditory environment*, J. Acoust. Soc. Am., **131**, 4, 2948–2957, doi: 10.1121/1.3687448.
24. Röder B., Teder-Sälejärvi W., Sterr A., Rösler F., Hillyard S.A., Neville H.J. (1999), *Improved auditory spatial tuning in blind humans*, Nature, Jul 8, **400**, 6740, 162–166.
25. Shinn-Cunningham B.G., Streeter T., Gyss J.-F. (2005), *Perceptual plasticity in spatial auditory displays*, Proceedings of the 2001 International Conference on Auditory Display, Espoo, Finland, July 29 – August 1, 2001.
26. Strelnikov K., Rosito M., Barone P. (2011), *Effect of Audiovisual Training on Monaural Spatial Hearing in Horizontal Plane*, PLoS ONE, **6**, 3.

27. Wenzel E.M. (2001), *Effect of increasing system latency on localization of virtual sounds with short and long duration*, Proceedings of the International Conference on Auditory Display, Espoo, Finland, July 29 – August 1, 2001, 185–190.

28. Wersényi Gy. (2012), *Virtual Localization by Blind Persons*, JAES, **60**, 7/8, 568–579.

29. Zahorik P., Bangayan P., Sundareswaran V., Wang K., Tam C. (2006), *Perceptual recalibration in human sound localization: Learning to remediate front-back reversals*, J. Acoust. Soc. Am., **120**, 343–359, doi: 10.1121/1.2208429.

30. Zwiers M.P., Van Opstal A.J., Cruysberg J.R. (2001), *Two-dimensional sound-localization behavior of early-blind humans*, Exp. Brain Res., **140**, 2, 206–222.

31. MIT HRTF Database. Available at: http://sound.media.mit.edu/resources/KEMAR.html