Influence of vision on short-term sound localization training with non-individualized HRTF

Tifanie Bouchara, Tristan-Gaël Bara, Pierre-Louis Weiss, Alma Guilbert

INFLUENCE OF VISION ON SHORT-TERM SOUND LOCALIZATION TRAINING WITH NON-INDIVIDUALIZED HRTF

Tifanie Bouchara[1] **Tristan-Gaël Bara**[1 2]
Pierre-Louis Weiss[2] **Alma Guilbert**[2]
[1] CEDRIC (EA4626), CNAM, HeSam Université ,75003, Paris, France
[2] VAC Laboratory (EA 7326), Université Paris Descartes, 92774 Boulogne-Billancourt, Paris, France
tifanie.bouchara@cnam.fr, alma.guilbert@parisdescartes.fr

ABSTRACT

Previous studies have demonstrated that it is possible for humans to adapt to new HRTF, non-individualized or altered, in a short time period through various training programs going from simple sound exposure to active learning. While all training programs are based on a bimodal coupling (audio-vision or audio-proprioception), they are rarely based on a trimodal one. Our study compares two versions of active trainings: an audio-proprioceptive one and an audio-visuo-proprioceptive one, in order to explicit the role of vision in short-term audio localization training when action and proprioception are already involved. Results from an experimental between-subjects design study, with 27 participants trained on three different program conditions with or without vision, reveal that vision seems to have no or very little influence on rapid audio-proprioceptive trainings and HRTF adaptation. This study could also help a better utilization of 3D-audio for neurological or psychiatric rehabilitation program.

1. INTRODUCTION

Auditory spatial perception and sound source localization rely on several auditory cues (binaural and spectral cues [1]) that allow us to estimate the position of a given sound source. They are contained in the Head Related Transfer Function (HRTF) simulating the transformations caused by the head, the pinna and the torso, for a sound given position. HRTF thus depends on the anatomical features of the listener, and are deeply individualized. Measuring them for each individual user of a VR system is a very long and expensive process. Therefore, it is frequent to use a set of HRTF, measured on another individual or even computationally generated.

Several studies have demonstrated that it is possible to adapt our auditory localization system to new or altered HRTF, for example naturally due to hair cut change and morphologic evolution [2], or with the help of training programs. Three types of training programs, i.e. sound exposition, feedback, and active learning, have been developed. The concept of sound exposition training is based on HRTF modification by artificially altering the outer ear. For example in [3], participants wore earmolds for as long as sixty days. The results showed a significant improvement in sound localization. However, the improvement did not lead to equal performance as those measured with individual HRTF. Feedbacks have also been used to study their effect on the adaptation to new HRTF. [4] used visual feedbacks to provide the correct position of the sound source, after each estimation of its localization. According to [5], feedbacks on the estimation during a localization task improved the performance in a greater measure than a simple exposition to the sound. Lastly, some studies also investigated the effect of active learning. [6] designed a method involving procedural and active learning using visual feedbacks. Once the participants pointed towards the perceived sound, they received a visual feedback at the position of the sound source. They had next to correct their estimation by pointing in the direction of the visual target. Then, to associate the visual and the auditive modalities, both targets were presented at the same time, and had to be pointed again. It is also possible to improve the performance with an active comparison of sounds [7, 8]. Finally, it has been proved that it is possible to shorten the adaptation phase using an active and implicit learning task as suggested by Parseihian and colleagues [9]. Carried out on 3 consecutive days, their training program consisted in a mini sonified version of a hotandcold game where blindfolded participants actively explore the sphere around them to search for invisible targets using a position-tracked ball held in their hand. This game-like task has the advantage to foster the immersion in the audio-virtual environment.

All these training methods are based on multimodal learning. According to [10], these kinds of learning methods are more effective than unimodal learning methods. Furthermore, multimodal learning influences unisensory perception. Indeed, it has been shown that presenting congruent auditory and visual stimuli during the learning stage leads to a greater improvement in visual performance than a visual-only training [11, 12]. Other studies demonstrated that congruent auditory and visual stimuli could also benefit for auditory performance [13]

doi:10.25836/sasp.2019.04

While all training programs are based on a bimodal coupling (audiovision [14] or audio proprioception [9]), they are rarely based on a trimodal one. It is suprising as many studies have shown that a multisensory learning could be more efficient than a unisensory learning (for a review see [10]). The study [9] has proved that it is possible to rapidly adapt to new HRTF without vision, but it is unclear if vision can reinforce or shorten this adaptation effect. We carried out an experiment presented in the next section to verify our hypothesis that seeing the sound source, during a training program involving implicit learning through an active exploration of the peripersonal sphere, leads to better localization progress than without vision.

2. EXPERIMENTATION

In contrast to earlier studies [9, 15], in which participants were systematically blinded, the aim of our study is to evaluate the influence of vision on sound localization training programs.

2.1 Procedure

2.1.1 Tasks and stimuli

	Day		
	1	2	3
Pre-localization test	L1		
Training task	T1	T2	T3
Post localization test	L2	L3	L4

Table 1. Task presentation across sessions.

As in [9], participants carried out two different tasks: one training task for the adaptation to non-individualized HRTF and one localization task for the assessment of this adaptation. All participants received a training session of 12 minutes during 3 consecutive days and also had to perform 4 sound localization tests: one before the experiment (L1) and one after each training session (L2 to L4).

The adaptation task was similar to the mini-game used in [9] and [15]. Participants had to freely scan the space around them with their hand-held position-tracked Vive controller in order to find an animal sound hidden around them. Target positions were randomly chosen in the frontal hemisphere. The controller-to-target angular distance is sonified through the alternate speed from a white and a pink noise such as the delay between each burst decreased from 3.0 s to 0.05 s with the angular distance (3.0 s meant the target was at the opposite direction). When the target position was reached, the search feedback sound was replaced by a random animal sound. When applicable (groups G_{AP} and G_{AVP}, see experimental design section), the feedback sound and the animal sounds were spatialized through binaural audio at the controller position. Animal sounds were taken from various free sample databases. Participants were asked to find as much as animal sounds they can during the 12 minutes of each adaptation session (T1 to T3, one per day).

Group	Modalities (including proprioception)	HRTF spatialization
Gc	Audio	none
Gap	Audio	non-individual
Gavp	Audio + Vision	non-individual

Table 2. Training task conditions per group.

In the localization tasks (L1 to L4), participants had to report the perceived position of a static spatialized sound sample by pointing with the hand-held controller and validating the direction with the trigger of the controller. As in [9], the stimulus consisted of a train of three 40 ms Gaussian broadband noise bursts (50–20 000 Hz) separated by 30 ms of silence. Each localization test was composed of 2 blocks of 33 trials testing localization performance for 11 azimuths $\{-90°, -72°, -54°, -36°, -18°, 0°, +18°, +36°, +54°, +72°, +90°\}$ × 3 elevations $\{-30°, 0°, +30°\}$. At the end of each new trial, participants first had to point a target presented visually (green object) at a position of $0°$ az., $0°$ el. so participants were always oriented towards the frontal direction at the beginning of a trial. In each block, trials were randomly presented. The mean duration of this task was 3 min per block. Participants were allowed to take a break of 3 min between blocks and tasks. The experiment lasted almost one hour the first day, then 30 minutes on days 2 and 3, for each participant.

2.1.2 Experimental design

We used a between-subjects design where participants were randomly assigned to 3 experimental groups whose conditions are summarized in Table 2.
G_C was a control group, i.e. participants of that group received a training session but it was impossible to gain any HRTF adaptation during that phase because sounds were displayed in mono without any binaural spatialization effect. Participants were still active but any gain in performance from session L1 to L4 could only be attributed to procedural learning effect (due to task repetition) and not to localization or binaural learning. In this condition, searching for hidden target was still possible as it relies on sonification processes and not in sound localization.
G_{AP} received an audio-proprioceptive training as exposed in [9, 15]. The animal sounds and the feedback sound indicating the angular distance to the target were spatialized with binaural. No visual information was provided.
G_{AVP} received the same task as in G_{AP} but, a visual representation of a sphere was also displayed at the hand position during all training sessions (audio-visuo-proprioceptive situation).

2.2 Hypotheses

We hypothesized that: H1) training programs would lead to an auditory adaptation when using implicit active learning and HRTF presentation ; H2) combining all modalities in the audio-visuo-proprioceptive program would optimize

doi:10.25836/sasp.2019.04

the adaptation, inducing better performance and a longer remaining effect than an audio-proprioceptive program.

2.3 Participants

Twenty-seven volunteers (age: 18-43 years (mean 23.5, St. 5.7); 24 females; 25 right-handed), essentially students from undergraduate programs of Paris Descartes University, participated in this between-subject designed study. They were randomly assigned to the experimental groups, finally composed as such: G_C (6 participants, 6 women, all right-handed, mean age 23 (ST 4.8)), G_{AP} (11 participants, 2 men, all right-handed, mean age 23.7 (ST 4.9)), and G_{AVP} (10 participants, 2 men, 8 right-handed, mean age 23.7 (ST 7.3)).

Participants were tested individually in the same isolated listening room, seated in a swivel chair. All participants reported normal or corrected to normal vision and normal hearing. All participants had an audiometric test before the experiment, verifying normal audiometric thresholds (less than 20 dB HL) at octave frequencies between 250 and 8000 Hz, and no history of hearing difficulties. Subjects were naive to the purpose of the experiment and the sets of spatial positions selected for the experiment. Informed consent was obtained from all participants. This experiment was approved by the Paris Descartes University Ethical Committee (CER) (authorization number: $N°2019-24$).

2.4 Apparatus

The experiment was conducted in a controlled lab environment. The audio-virtual environment was developed under Unity with Steam VR, and was rendered using a HTC Vive as a head- and hand-tracker. 3D audio spatialization is obtained through Steam Audio's non-individualized built-in HRTF. When applicable, 3D visual information is displayed directly on the Vive screen. The computer was composed of two-intel core i7-4790K as CPU, two GeForce GTX 980 as GPU, 16GhZ of RAM and a MSI Z97 Gaming 5 motherboard. Open circum-aural reference headphones (Sennheiser HD 380 Pro) were used without any headphone compensation.

3. RESULTS

3.1 Localization task

3.1.1 Dependant variables and analyses methodology

Target and response azimuths and elevations were logged for each trial during the different localization tests. These dependent variables were converted to the interaural polar coordinate system (lateral and polar angle), initially presented in [16], to analyze the type of error (precision, front/back confusion and up/down confusion) as explained in [9] and [15]. The lateral angle ($-90° \leqslant \alpha \leqslant 90°$) was calculated from the median plane to the represented vector ; the polar angle ($-90° \leqslant \alpha \leqslant 270°$) indicates the rotation around the interaural axis, with $0°$ being front. Then all

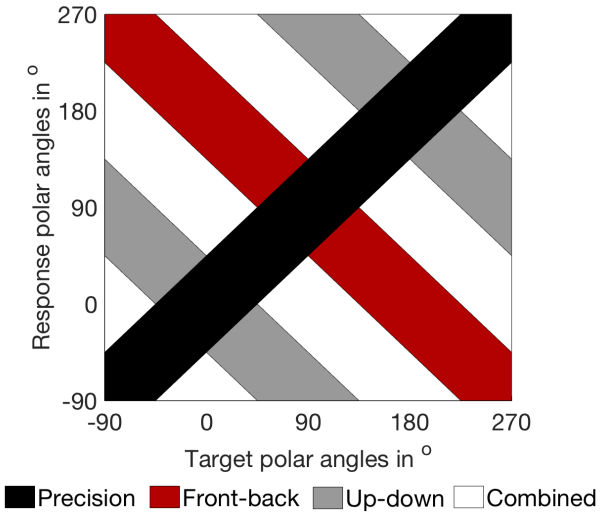

Figure 1. Definition of the four different error type zones according to [9].

error types were determined according to [9] and [15] using the different zones of scatter plot response versus target polar angle as presented in Figure 1.

As the distributions of absolute lateral and polar errors were not normal due to front-back confusions [9], we used the median instead of the mean to analyse the lateral and polar errors. Table 3 reports the average median of localization errors in lateral and polar angles, and the average percentages of front-back confusions for the different groups and localization tests.

A Kruskal-Wallis (non-parametric) ANOVA was carried out on the three dependent variables, lateral and polar errors and percentages of front-back inversion to determine the effect of the group (inter-group comparisons) before the training (localization test L1) and after the training (test L4). A Friedman (non-parametric) ANOVA was carried out on the same three dependent variables to evaluate the adaptation effect, i.e. to determine the effect of the sessions in each group. Wilcoxon tests were carried out to find if there is a significant difference between sessions L1 and L4. Finally, in order to determine which participant showed a learning effect, we carried out intra-subject analyses: Mann-Whitney tests were done on our three dependent variables to compare session L1 to L4. A significance threshold of .05 (one-tailed alpha level) was adopted for all statistical analyses.

3.1.2 Lateral Error

The mean of absolute lateral error medians is shown in Figure 2 for each group over the course of the 4 localization tests (L1 to L4).

Our three groups were not significantly different in session L1 ($H(dl{=}2, N{=}27){=}0.41, p{=}.41$) nor in session L4 ($H(dl{=}2, N{=}27){=}.48, p{=}.39$) for lateral errors.

A significant improvement across sessions was found for G_{AP} ($F(dl{=}2, N{=}11){=}10.2, p{=}.0085$). Participants were better in session L4 than in

		Localization test			
		L1	L2	L3	L4
G_C	Lateral Error Medians	18.3 (3.8)	15.8 (3.2)	16.0 (3.8)	17.1 (5.7)
	Polar Error Medians	88.6 (65.7)	82.6 (57.4)	86.5 (56.1)	76.1 (54.1)
	Percent of Front-Back Error	28.3 (20.9)	21.4 (18.7)	23.5 21.8)	21.0 (23.2)
G_{AP}	Lateral Error Medians	19.9 (5.0)	19.5 (4.1)	17.6 (5.6)	16.0 (3.8)
	Polar Error Medians	58.0 (27.9)	66.7 (41.6)	64.4 (39.6)	60.2 (45.4)
	Percent of Front-Back Error	21.8 (14.5)	19.0 (16.4)	20.2 (19.5)	17.8 (19.9)
G_{AVP}	Lateral Error Medians	18.8 (7.4)	17.9 (7.4)	17.9 (7.4)	17.3 (5.2)
	Polar Error Medians	107.7 (60.6)	103.3 (64.0)	95.7 (59.5)	102.5 (65.3)
	Percent of Front-Back Error	29.7 (24.4)	29.1 (24.8)	24.6 (22.6)	26.1 (23.0)

Table 3. Mean of lateral error medians (in °), mean of polar error medians (in °), and percentages of front-back errors across session and groups. Numbers in brackets indicate standard deviations.

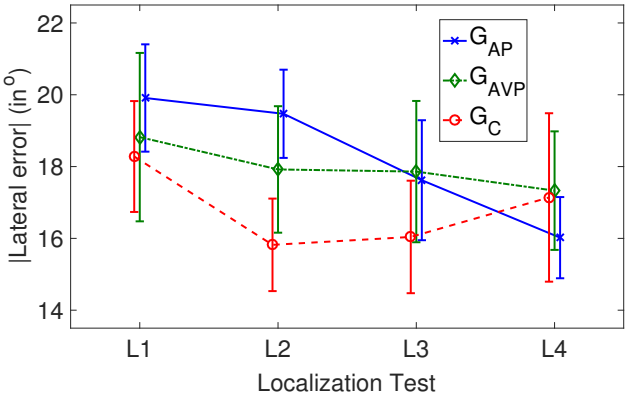

Figure 2. Mean absolute lateral error medians during the four localization tests for each group (control without HRTF presentation, audio-proprioceptive and audio-visuo-proprioceptive). Vertical bars indicate standard error of the mean.

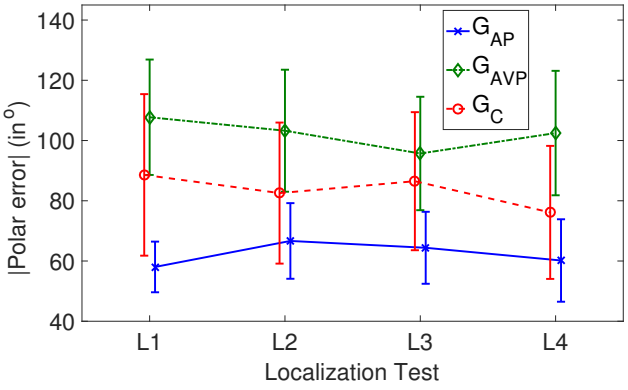

Figure 3. Mean absolute polar error medians during the four localization tests for each group (control without HRTF presentation, audio-proprioceptive and audio-visuo-proprioceptive). Vertical bars indicate standard error of the mean.

session L1 (Z=2.67, p=.0038). No significant improvement across the sessions was found for G_{AVP} ($F(dl$=2, N=10)=0.72, p=.43) and G_C ($F(dl$=2, N=6)=2.69, p=.22). Again, no improvement between session 1 and session 4 was found for G_{AVP} (Z=0.76, p=.22) and G_C (Z=0.52, p=.30).

Intra-subject analyses revealed that 5/11 participants from G_{AP}, 5/10 from G_{AVP}, and 3/6 from G_C were or tended to be better for session L4 than for session L1.

3.1.3 Polar Error

For both sessions L1 and L4, our three groups were not significantly different in polar errors (L1: $H(dl$=2, N=27)=2.74, p=.13; L4: $H(dl$=2, N=27)=2.30, p=.16). However there was a statistical tendency to a difference between G_{AP} and G_{AVP} in both session L1 before learning (U=32, Z= − 1.58, p=.057) and session L4 after learning (U=35, Z= − 1.37, p=.087).

No significant effect of the session was found neither for G_{AP} ($F(dl$=2, N=11)=0.054, p=.5) nor for G_C ($F(dl$=2, N=6)=3.2, p=.18). Only a statistical tendency for an effect of session was found for the group G_{AVP} ($F(dl$=2, N=10)=5.88, p=.059). However this was not due to learning as the difference between L1 and L4 for that group G_{AVP} was not significant (Z=0.005, p=.48).

Intra-subject analyses of absolute polar errors revealed that 5/11 participants from G_{AP}, 1/10 from G_{AVP} and 2/6 from G_C were or tended to be better for session L4 than for session L1.

3.1.4 Front-back confusion

Our three groups were not significantly different in percentage of front-back inversion in session L1 ($H(dl$=2, N=27)=0.35, p=.42) nor for session L4 ($H(dl$=2, N=27)=0.42, p=.40).

No significant improvement across sessions was found for none of the three groups G_{AP} ($F(dl$=2, N=11)=1.78, p=.31), G_{AVP} ($F(dl$=2, N=10)=2.6, p=.23) and G_C ($F(dl$=2, N=6)=3.86, p=.28). Only a statistical tendency between session L1 to L4 was found for the participants of group G_{AP} (Z=1.42, p=.077) with a trend to less front-back confusion errors at the end of the training.

doi:10.25836/sasp.2019.04

3.2 Adaptation task

No measure of localization abilities was done in this task. However, the number of sound sources found during this task can provide indications on the task difficulty. First, a comparison between the three trainings (T1 to T3) was carried out to underline the progress across trainings thanks to a Friedmann (non-parametric) ANOVA and Wilcoxon tests. Then, a comparison between our three groups was made thanks to a Kruskal-Wallis (non-parametric) ANOVA in order to determine if the use of HRTF or vision could make the task easier. Table 4 reports the average of completed trials per group, i.e. the number of animals found, during each session of 12 min.

The number of sound sources found was different across the sessions ($F(dl{=}2, N{=}27){=}29.7, p{<}.0001$). Participants performed better during T2 than during T1 ($Z{=}3.34, p{=}.0004$) and during T3 than during T2 ($Z{=}3.1, p{=}.001$). Performance improved across the sessions.

Our three groups did not show significant differences during any of the training ($T1{:}H(dl{=}2, N{=}27){=}0.21, p{=}.45$; $T2{:}H(dl{=}2, N{=}27){=}0.93, p{=}.31$; $T3{:}H(dl{=}2, N{=}27){=}0.075, p{=}.48$). The use of HRTF or vision did not seem to make the task easier.

	Training session		
	T1	T2	T3
G_C	13.2 (1.1)	15.0 (2.7)	20.2 (1.4)
G_{AP}	14.3 (8.1)	17.7 (7.2)	19.8 (7.0)
G_{AVP}	12.4 (6.1)	18.8 (6.9)	20.5 (5.4)

Table 4. Mean number of animals found per session (standard deviation) across session and groups.

4. DISCUSSION

The only significant improvement of localization performance across the sessions was observed for the audio-proprioceptive group G_{AP}: a significant learning effect was underlined on lateral errors between before and after the program. That means that our training program based on an active and implicit audio-proprioceptive learning task is efficient for non-individualized HRTF adaptation, as shown by [9]. No significant improvement was observed for the audio-visuo-proprioceptive group G_{AVP} and for the control group G_C. Vision did not lead to better HRTF localization performance in our study, suggesting that the program without vision could be better than the one with vision. This does not support our hypothesis of a better localization progress with vision than without.

Nevertheless, this result could be explained by some particularities of multisensory integration. As seen previously, multisensory learning could be more efficient than a unisensory learning on the perception of a particular modality [11–13]. Spatial and temporal coincidences facilitate multisensory integration [17]. Hypotheses state that discrepancies are resolved in favor of the more appropriate modality [14]. Although hearing is predominant for temporal perception, vision is more precise for spatial judgments. However, in our experiment, although the visual information is spatially and temporally congruent with the auditory one, vision did not seem to improve the learning of the spatial positions of the new HRTF. This absence of effect could be explained by the fact that multisensory integration depends on some other conditions. One of the main principles underlying multisensory integration is the inverse effectiveness rule [17]. According to this principle, the efficiency of the integration depends on the nature of stimuli: the more unimodal stimuli are ambiguous and weak, the more another modality is used, even if the information from this modality is not apparently related to the task [18]. In our adaptation task, vision was informative to localize the hand position in space. However, the visual information did not inform of the distance from the target, contrary to the auditory information. The auditory information may have appeared sufficient to come up with a robust estimate and realize the task and, thus, information from several modalities was not combined, such as suggested by the probabilistic model of [19], and visual information could even be distracting for the participants. One way to provoke more multisensory integration would be to give less information to the auditory modality and more to the visual one. We could imagine, in a future study, to give the information of distance from the target also to the visual modality in order to show if we could improve the visual contribution and, thus, improve the learning of new HRTF.

However, in our study, no difference between the three groups of participants can be observed at the end of the third and last training session. This could be due to a sample of participants being too small to show statistical significance. This could also be due to an insufficient number of learning sessions. Indeed, the audio-proprioceptive group GAP especially improved after the third training session. Therefore, we could hypothesize that the adaptation will be greater after a fourth session or more. This is coherent with the results of recent studies [15] who have shown a continuing improvement over a program of 10 weeks, one session per week. Another explanation would be that groups may be heterogeneous in terms of matching between the individual participant HRTF and the non-individualized generic HRTF. Indeed, the adaptation slope depends on the compatibility between the HRTF of the listener and the HRTF to be learnt [9] and, in each of our groups, some participants were able to improve their localization performance whereas others were not. Moreover, the audio-proprioceptive group G_{AP} and the audio-visuo-proprioceptive group G_{AVP} tended to be different in polar errors before learning. This suggests that our two groups may have had differences in terms of matching with the non-individualized generic HRTF before the learning. Another experiment with more participants and a preselection step to select compatible HRTF for all of these participants is finally required to

really assess if vision can or not improve adaptation to HRTF.

Another point that needs to be highlighted is that three participants from the control group improved across the sessions on the lateral and/or on the polar errors. Yet, this group was not exposed to any HRTF during the training. So, the statistical improvements can only be explained to a familiarization with the material and the task. This definitely is an argument to encourage further HRTF training studies to always compare with a control group performing the task in mono instead of comparing with a group receiving no training at all.

Finally, our study and future ones could improve the use of virtual reality and 3D audio as a rehabilitation strategy for specific listeners suffering from neurological disorders, such as spatial cognition disorders. For example, our program could be adapted for unilateral spatial neglect, which is a common neurological disorder in which patients have difficulties to pay attention to the contralesional side of space in vision, but also in other sensory modalities such as hearing [20].

5. REFERENCES

[1] J. Blauert, *Spatial Hearing, The Psychophysics of Human Sound Localization*. MIT Press, Oct. 1996.

[2] J. P. Rauschecker, "Auditory cortical plasticity: a comparison with other sensory systems.," *Trends neurosci.*, vol. 22, pp. 74–80, feb 1999.

[3] S. Carlile, K. Balachandar, and H. Kelly, "Accommodating to new ears: The effects of sensory and sensory-motor feedback," *J. Acoust. Soc. Am.*, vol. 135, pp. 2002–2011, apr 2014.

[4] 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.*, vol. 103, pp. 3656–3666, jun 1998.

[5] K. Strelnikov, M. Rosito, and P. Barone, "Effect of Audiovisual Training on Monaural Spatial Hearing in Horizontal Plane," *PLOS ONE*, vol. 6, pp. 1–9, 03 2011.

[6] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Atten. Percept. Psycho.*, vol. 72, pp. 454–469, Feb 2010.

[7] C. Mendonça, G. Campos, P. Dias, J. Vieira, J. P. Ferreira, and J. A. Santos, "On the improvement of localization accuracy with non-individualized HRTF-based sounds," *J. Audio Eng. Soc.*, vol. 60, pp. 821–830, Oct. 2012.

[8] C. Mendonça, G. Campos, P. Dias, and J. A. Santos, "Learning auditory space: Generalization and long-term effects," *PLOS ONE*, vol. 8, 10 2013.

[9] G. Parseihian and B. F. G. Katz, "Rapid head-related transfer function adaptation using a virtual auditory environment," *J. Acoust. Soc. Am.*, vol. 131, no. 4, pp. 2948–2957, 2012.

[10] L. Shams and A. R. Seitz, "Benefits of multisensory learning," *Trends Cogn. Sci.*, vol. 12, no. 11, pp. 411–417, 2008.

[11] A. R. Seitz, R. Kim, and L. Shams, "Sound Facilitates Visual Learning," *Curr. Biol.*, vol. 16, pp. 1422–1427, jul 2006.

[12] R. S. Kim, A. R. Seitz, and L. Shams, "Benefits of stimulus congruency for multisensory facilitation of visual learning," *PLOS ONE*, vol. 3, pp. 1–5, 01 2008.

[13] K. von Kriegstein and A.-L. Giraud, "Implicit multisensory associations influence voice recognition," *PLOS Biology*, vol. 4, pp. 1–12, 09 2006.

[14] C. Mendonça, "A review on auditory space adaptations to altered head-related cues," *Front. Neurosci.*, vol. 8, p. 219, 2014.

[15] P. Stitt, L. Picinali, and B. F. Katz, "Auditory Accommodation to Poorly Matched Non-Individual Spectral Localization Cues Through Active Learning," *Sci. Rep.*, vol. 9, no. 1, 2019.

[16] M. Morimoto and H. Aokata, "Localization cues of sound sources in the upper hemisphere.," *J. Acoust. Soc. Japan (E)*, vol. 5, no. 3, pp. 165–173, 1984.

[17] B. E. Stein and M. A. Meredith, *The merging of the senses*. Cambridge , MA, The MIT Press, Jan. 1993.

[18] E. B. Stein, N. London, K. L. Wilkinson, and P. Donald, "Enhancement of perceived visual intensity by auditory stimuli: A psychophysical analysis," *J. Cogn. Neurosci.*, vol. 8, pp. 497–506, 11 1996.

[19] M. O. Ernst and H. H. Bülthoff, "Merging the senses into a robust percept," *Trends Cogn. Sci.*, vol. 8, no. 4, pp. 162 – 169, 2004.

[20] A. Guilbert, S. Clément, L. Senouci, S. Pontzeele, Y. Martin, and C. Moroni, "Auditory lateralisation deficits in neglect patients," *Neuropsychologia*, vol. 85, pp. 177 – 183, 2016.

 doi:10.25836/sasp.2019.04