TRAINING SPATIAL HEARING SKILLS IN VIRTUAL REALITY THROUGH A SOUND-REACHING TASK

Chiara Valzolgher[1*] **Sara Capra**[3] **Kevin Sum**[2]
Francesco Pavani[134] **Lorenzo Picinali**[2]
[1] 1. Center for Mind/Brain Sciences (CIMeC), University of Trento, Italy
2. Audio Experience Design (www.axdesign.co.uk), Imperial College London, UK
3. Department of Psychology and Cognitive Sciences (DiPSCo), University of Trento, Italy
4. Centro Interuniversitario di Ricerca "Cognizione, Linguaggio e Sordità" (CIRCLeS), Trento, Italy

ABSTRACT[*]

Sound localization is crucial for interacting with the surrounding world. This ability can be learned across time and improved by multisensory and motor cues. In the last decade, studying the contributions of multisensory and motor cues has been facilitated by the increased adoption of virtual reality (VR). In a recent study, sound localization had been trained through a task where the visual stimuli were rendered through a VR headset, and the auditory ones through a loudspeaker moved around by the experimenter. Physically reaching to sound sources reduced sound localization errors faster and to a greater extent if compared to naming sources' positions. Interestingly, training efficacy extended also to hearing-impaired people. Yet, this approach is unfeasible for rehabilitation at home. Fully-virtual approaches have been used to study spatial hearing learning processes, performing headphones-rendered acoustic simulations. In the present study, we investigate whether the effects of our reaching-based training can be observed when taking advantage of such simulations, showing that the improvement is comparable between the full-VR and blended VR conditions. This validates the use of training paradigms that are completely based on portable equipment and don't require an external operator, opening new perspectives in the field of remote rehabilitation.

[*]***Corresponding author**: chaira.valzolgher@unitn.it.*

Keywords: *sound localization, virtual reality, reaching to sounds, virtual sounds, hearing rehabilitation.*

1. INTRODUCTION

Identifying where the sounds come from is crucial for interacting with the surrounding world. The so-called sound localization ability can be learned across time and improved by multisensory and motor cues [1]. In the last decade, virtual reality technologies have played a crucial role in studying of sound localization learning processes. In a recent study, sound localization was trained through a task where the visual stimuli were rendered through a VR headset, and the auditory ones through a loudspeaker moved around by the experimenter [2, 3]. Participants were instructed to physically reach toward sound sources. The findings showed that this type of interaction with sound sources reduced sound localization errors faster and to a greater extent if compared to a control condition in which participants were asked to name sources' labels to identify them. Interestingly, training efficacy generalized to a different sound localization task and extended also to hearing-impaired people [4-7]. Yet, this approach is unfeasible for rehabilitation at home, and even in the research practice, it requires a great effort and time on the part of the experimenter to move the speaker. Fully-virtual approaches, which rely on the use of binaural spatialization techniques for the audio rendering and delivery through a pair of headphones, have been used to study spatial hearing learning processes [8-10]. The main difference between these (headphone-based) and a loudspeaker-based approach is indeed in the overall goal of the simulation; the latter aims at re-creating a simulated sound field in an area where

the listener is located, and the first in reproducing two distinct signals at the ears of the listener, embedding in these the localization cues that then are used to determine, for example, the position of the various sound sources. While headphone-based approaches are certainly more accessible and portable, they rely on simplifications when rendering the audio scenes; for example, ideally, they'd need to be personalized for every individual listener [8], but in reality, they rely on generalizations, which could result in lower accuracy of the rendered scene.
In the present study, we investigated whether the effects of the training based on reaching toward sound sources can be observed by implementing a similar paradigm by taking advantage of a fully virtual reality approach.

2. MATERIAL AND METHOD

The present work focuses on part of the data collected in two previous studies [2, 11] (ethic protocols 2018-018 and 2022-009). The original study involving 28 normal-hearing participants [2] was run in 2020, while the other one, involving other 28 normal-hearing participants [11] in 2023. Sample sizes and the rationales of the two studies were therefore determined by the respective research questions.

2.1. Apparatus and Stimuli

In both experimental scenarios [2, 11], participants were asked to wear a Head-Mounted Display (HMD, HTC Vive, and Meta Quest 2 respectively) and be immersed in a square room, developed with platform Unity3D (Unity Technologies, San Francisco, CA), similar in size to the real ones (4*3 meters and 2*2,5 meters). The room was empty, with a door behind participants and designed with exposed bricks. Participants saw 17 speakers distributed in a semicircle in front of them at the ears level, at 55 cm from participants' head, and spaced about ±80° of visual angle. Above each speaker, a numerical label from 1 to 17, was located. These labels change randomly trial by trial.
In the experiment 1 [2], white noise sounds were emitted by a real speaker moved by the experimenter around participants [3-7]. The experimenter was guided to reach each of the predetermined positions by following instruction provided on a monitor. In the experiment 2 [11] white noise sounds were instead transmitted using headphones and sound spatialization was performed using the Unity integration of a convolution-based algorithm named 3D Tune-In Toolkit [12]. Sounds were spatialized using the HRTF of a KEMAR dummy head mannequin from the SADIE database [13], and Interaural Time Differences (ITDs) were customized for each individual user according to their head circumference.

2.2. Stimuli and Procedure

In both studies, participants were invited to sit on a chair and were instructed to wear the head mounted display, and hold the controller in their right hand. All participants were instructed to perform a single sounds localization task following different instructions as a function of their group. Specifically, in this article, we focused on 2 groups of 14 people each for both experiments (for a total of 28 participants in each experiment). The Reaching group (the experimental group) was instructed to localize sound by reaching the source with the controller held in their right hand. The Naming group (the control group) was instructed to localize sound sources by naming the numerical label located over the speakers. Both groups received audio-visual feedback about their performance: if they reached or named the correct speaker, the sounds stopped. While if they reached or named an incorrect speaker, the sounds did not stop and the correct speaker started to flash until they provided the correct response.
All participants performed one block (B1) while listening in normal hearing condition (binaural) and two successive blocks (B2 and B3) in altered listening condition. To simulate altered listening conditions, participants who performed the experiment 1 were monaurally plugged by using an earplug (3M PP 01 002). Differently, in the experiment 2 participants were exposed to simulation of a mild-to-moderate monolateral hearing loss, which was obtained again using the 3D Tune-In Toolkit. It is clear that there are notable differences between using an earplug and simulating a hearing loss, especially when looking at non-linear effects. Nevertheless, the use of the earplug aimed at replicating the effect of a monolateral hearing loss, which could be more accurately simulated in the virtual domain (yet another potential advantage of this approach). Note that there was a difference in the number of trials for each block: in the experiment 1, blocks comprised 51 (17*3) trials, while in the experiment 2 blocks comprised 68 (17*4) trials. The reason is that we did not plan to test this comparison when we designed experiment 2.

2.3. Analysis

To document the performance, we measured the absolute error along the horizontal dimension obtained by calculating the discrepancy between the position of the speaker and participants responses for each trial respectively in absolute or signed values. In the analysis we considered log transformed absolute error in order to correct the skewness of distributions.

3. RESULTS

3.1 Sound localization: effect of altering listening experience

To assess the effect of altering listening experience, we entered absolute errors into an ANOVA with one within factor block (normal vs altered hearing) and two between factors: interaction with sounds group (reaching, naming) and type of sound group (real, virtual). We found a main effect of block indicating that the errors increased during the altered as compared to the normal hearing condition ($F(1,52) = 212.00$, $p < 0.001$) irrespectively of both interaction with sounds and type of sound (all *ps* > 0.26) (see Table 1).

Table 1. Mean absolute errors as a function of block, interactions with sounds and types of sound.

	Block		
	1	*2*	*3*
Reaching			
Real (exp 1)	2.17	12.23	8.63
Virtual (exp 2)	3.17	12.71	5.96
Naming			
Real (exp 1)	1.93	11.37	12.08
Virtual (exp 2)	2.95	12.08	9.24

3.2. Sound localization during altered listening

To assess the effect of repeating altered listening experience, we entered errors into an ANOVA with one within factor block (first altered block B2, second altered block B3) and two between factors: interaction with sounds group (reaching, naming) and type of sound group (real, virtual). We found a main effect of block, indicating that the errors decreased in the B3 compared to B2 ($F(1,52) = 24.96$, $p < 0.001$). We also found a main effect of the type of sound indicating that the errors were overall lower when participants were exposed to altered virtual sounds (simulating unilateral hearing loss) as compared to real sounds (listened monaurally) ($F(1,52) = 4.13$, $p = 0.047$). We documented that reaching to sound sources was able to reduce error more than naming irrespectively of the type of the sound (interaction between interaction with sounds and block, $F(1,52) = 4.31$, p = 0.042). Finally, we found an interaction between the type of sound and block indicating that the amount of improvement (the difference between B2 and B3) was higher when participants were exposed to altered virtual sounds (simulating unilateral hearing loss) as compared to real sounds (listened monaurally) $F(1,52) = 6.71$, $p = 0.01$) (see Table 1 and Figure 1).

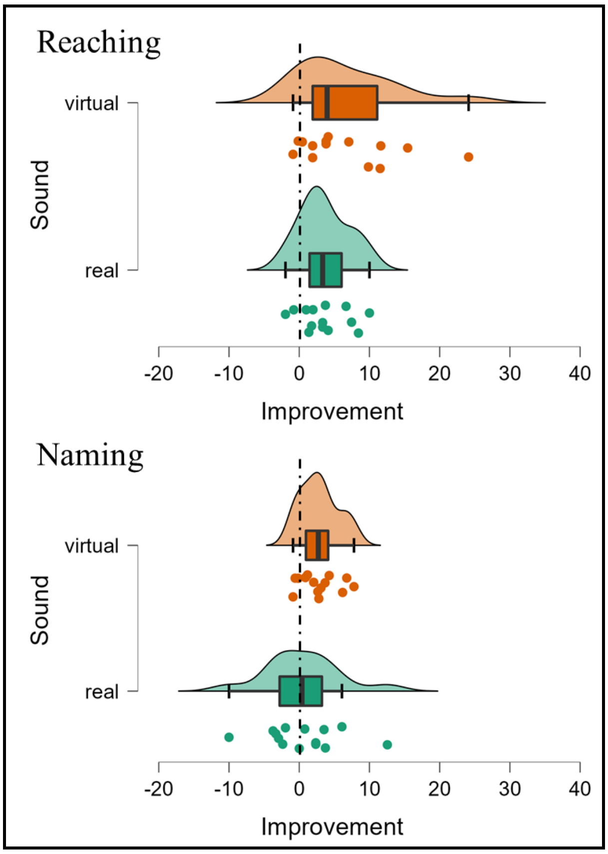

Figure 1. Differences between the absolute errors obtained in the block 2 (B2) and in the block 3 (B3) as a function of type of interaction with sounds (reaching, plot above; naming, plot below) and type of sound (virtual, red; real, green). Positive values indicate error reduction in the B3 as compared to the B2 (i.e. improvement).

4. DISCUSSION

This study demonstrated that the effect of a training paradigm based on reaching to sounds performed by using real sounds is comparable to the effect obtained by the same training paradigm performed by adopting virtual sound stimulations. This result contributed to demonstrating the effectiveness of reaching to sounds when training sound localization abilities and proved for the first time that its

efficacy remains even when adopting different types of auditory stimulation.
We also found that the amount of improvement was higher when participants were exposed to altered virtual sounds as compared to real sounds suggesting that training with virtual sounds may have led to better adaptation to altered hearing experience. However, this finding could be due to other differences between the two experiments. A factor which could have played a role in the improvement is the difference in the number of trials (51 vs 68 per block): seventeen more trials may have been decisive considering the total number of trials performed. Plus, the different way in which the hearing experience was altered could have an impact on the possibility of quickly adapting to altered auditory cues. The simulation of moderate unilateral hearing loss proposed in the experiment 2 may have given access to greater and richer acoustic information than using the ear plug as in the experiment 1.

5. CONCLUSIONS

In sum, these findings are sufficient to conclude that the reaching to sounds training paradigm clearly maintains its documented effectiveness, even when it was completely performed within acoustic virtual reality conditions. In this way, we validated the use of a training paradigm that is completely based on portable equipment and does not require an external operator. This example opens new perspectives in the field of remote rehabilitation and encourages research to study ecological and portable solutions, which can be important both to study perception and cognition in an ecological context [14] and to increase the possibility of accessing training protocols for people with hearing deficits.

6. ACKNOWLEDGMENTS

We are grateful to all the participants in the study and to the authors which collaborated in the experiment 1. The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: CV and FP were supported by a grant of the Velux Stiftung (n. 1439). FP was also financed by a grant from the Italian Ministry for Research and University (MUR, PRIN 20177894ZH).

7. REFERENCES

[1] Mendonça, C. (2014). A review on auditory space adaptations to altered head-related cues. *Frontiers in neuroscience*, 8, 219.

[2] Valzolgher, C., Verdelet, G., Salemme, R., Lombardi, L., Gaveau, V., Farné, A., & Pavani, F. (2020). Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues. *Neuropsychologia*, 149, 107665.

[3] Gaveau, V., Coudert, A., Salemme, R., Koun, E., Desoche, C., Truy, E., ... & Pavani, F. (2022). Benefits of active listening during 3D sound localization. *Experimental Brain Research*, 240(11), 2817-2833.

[4] Valzolgher, C., Todeschini, M., Verdelet, G., Gatel, J., Salemme, R., Gaveau, V., ... & Pavani, F. (2022). Adapting to altered auditory cues: Generalization from manual reaching to head pointing. *Plos one*, 17(4), e0263509.

[5] Valzolgher, C., Gatel, J., Bouzaid, S., Grenouillet, S., Todeschini, M., Verdelet, G., ... & Pavani, F. (2023). Reaching to sounds improves spatial hearing in bilateral cochlear implant users. *Ear and Hearing*, 44(1), 189-198.

[6] Valzolgher, C., Bouzaid, S., Grenouillet, S., Gatel, J., Ratenet, L., Murenu, F., ... & Pavani, F. (2023). Training spatial hearing in unilateral cochlear implant users through reaching to sounds in virtual reality. *European Archives of Oto-Rhino-Laryngology*, 1-12.

[7] Alzaher, M.; Valzolgher, C.; Verdelet, G.; Pavani, F.; Farnè, A.; Barone, P.; Marx, M. Audiovisual Training in Virtual Reality Improves Auditory Spatial Adaptation in Unilateral Hearing Loss Patients. *J. Clin. Med*. 2023, 12, 2357.

[8] Picinali, L., & Katz, B. F. (2022). System-to-User and User-to-System Adaptations in Binaural Audio. In Sonic Interactions in Virtual Environments (pp. 115-143). Cham: Springer International Publishing.

[9] Steadman, M. A., Kim, C., Lestang, J. H., Goodman, D. F., & Picinali, L. (2019). Short-term effects of sound localization training in virtual reality. *Scientific Reports*, 9(1), 1-17.

[10] Vickers, D., Salorio-Corbetto, M., Driver, S., Rocca, C., Levtov, Y., Sum, K., ... & Picinali, L. (2021). Involving Children and Teenagers With Bilateral Cochlear Implants in the Design of the BEARS (Both EARS) Virtual Reality Training Suite Improves Personalization. *Frontiers in Digital Health*, 3, 759723.

[11] Valzolgher, C., Capra, S., Sum, K., Pavani, F., & Picinali, L., under review. Spatial hearing training in acoustic virtual reality: a comparison between pointing, reaching and naming. *Scientific Reports.*

[12] Cuevas-Rodríguez, M., Picinali, L., González-Toledo, D., Garre, C., de la Rubia-Cuestas, E., Molina-Tanco, L., & Reyes-Lecuona, A. (2019). 3D Tune-In Toolkit: An open-source library for real-time binaural spatialisation. *PloS one*, 14(3), e0211899.

[13] Armstrong, C., Thresh, L., Murphy, D., & Kearney, G. (2018). A perceptual evaluation of individual and non-individual HRTFs: A case study of the SADIE II database. *Applied Sciences*, 8(11), 2029.

[14] Hartley, C. A. (2022). How do natural environments shape adaptive cognition across the lifespan? *Trends in Cognitive Sciences*.