Current Research in Behavioral Sciences 1 (2020) 100003

Contents lists available at ScienceDirect

Current Research in Behavioral Sciences

journal homepage: www.elsevier.com/locate/crbeha

The impact of a visual spatial frame on real sound-source localization in virtual reality

Chiara Valzolgher[a,b,*], Mariam Alzhaler[d], Elena Gessa[c], Michela Todeschini[c], Pauline Nieto[d], Gregoire Verdelet[a,g], Romeo Salemme[a,g], Valerie Gaveau[a,e], Mathieu Marx[h], Eric Truy[f], Pascal Barone[d], Alessandro Farnè[a,b,g], Francesco Pavani[a,b,c]

[a] Integrative, Multisensory, Perception, Action and Cognition Team (IMPACT), Lyon Neuroscience Research Center, France
[b] Center for Mind/Brain Sciences (CIMeC), University of Trento, Italy
[c] Department of Psychology and Cognitive Sciences (DiPSCo), University of Trento, Italy
[d] Centre de Recherche Cerveau & Cognition, Toulouse, France
[e] University of Lyon 1, France
[f] Hospices Civils de Lyon, Lyon, France
[g] Neuroimmersion, Lyon, France
[h] Hospices Civils, Toulouse, France

ARTICLE INFO

Keywords:
Spatial hearing
Virtual reality
Visual information
Active listening
Head movements tracking
Gaze tracking

ABSTRACT

Studies on audio-visual interactions in sound localization have primarily focused on the relations between the spatial position of sounds and their perceived visual source, as in the famous ventriloquist effect. Much less work has examined the effects on sound localization of seeing aspects of the visual environment. In this study, we took advantage of an innovative method for the study of spatial hearing – based on real sounds, virtual reality and real-time kinematic tracking – to examine the impact of a minimal visual spatial frame on sound localization. We tested sound localization in normal hearing participants (N=36) in two visual conditions: a uniform gray scene and a simple visual environment comprising only a grid. In both cases, no visual cues about the sound sources were provided. During and after sound emission, participants were free to move their head and eyes without restriction. We found that the presence of a visual spatial frame improved hand-pointing in elevation. In addition, it determined faster first-gaze movements to sounds. Our findings show that sound localization benefits from the presence of a minimal visual spatial frame and confirm the importance of combining kinematic tracking and virtual reality when aiming to reveal the multisensory and motor contributions to spatial-hearing abilities.

1. Introduction

In humans, as well as in other animals that can hear, the ability to localize sounds in space has evolved over the years within a multisensory environment. Under this ecological pressure, spatial hearing co-evolved with other sensory systems such as vision, which provides distal information about the environment (Heffner and Heffner, 1992, Heffner and Heffner, 2014). In addition, studies in animal and human models clearly showed that vision plays a critical role in the development of acoustic space perception (Knudsen and Knudsen, 1985, Hofman et al., 1998). When a listener is engaged in a sound localization task, there are at least two ways in which vision can contribute useful spatial information. First, vision can provide *direct* information about the auditory target, by revealing the position of the sound source in the environment (e.g., the listener hears and sees the bird tweeting on the tree). Second, vision can provide *indirect* information about the auditory targets, by revealing from which sector of space they may originate or by providing general information about the environmental spatial frame for encoding sound position (e.g., the listener cannot see the bird tweeting, but perceives the tree branches from which the stimulus originates).

The vast majority of studies that investigated audio-visual interactions in spatial hearing have been carried out in a context in which vision provides direct visual cues about sound position. In the typical experiment of this sort, the onset of the target sound is accompanied by a visual cue. When the visual information is veridical, listeners are more precise in sound localization compared to when no information is provided (Shelton and Searle, 1980, Tonelli et al., 2015). Instead, when the visual information is not veridical, visual-capture of sound position typically

* Corresponding author. Present address: Integrative, Multisensory, Perception, Action and Cognition Team (IMPACT), Centre de Recherche en Neurosciences de Lyon Inserm U1028 - CNRS UMR5292 Bâtiment Inserm 16 avenue Doyen Lépine 69676 BRON Cedex.
E-mail addresses: chiara.valzolgher@unitn.it, chiara.valzolgher@inserm.fr (C. Valzolgher).

https://doi.org/10.1016/j.crbeha.2020.100003
Received 12 August 2020; Received in revised form 19 October 2020; Accepted 10 November 2020
2666-5182/© 2020 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY license (http://creativecommons.org/licenses/by/4.0/)

emerges (the well-known 'ventriloquist effect') (Alais and Burr, 2004). For instance, Bolognini et al. (2007) demonstrated that veridical visual cues can enhance sound localization. Participants sat in front of a plastic semicircular apparatus which comprised eight loudspeakers hidden behind a curtain. They were required to verbally judge sound location, reading aloud labels marking the position of each speaker. Crucially, the auditory stimulus was either presented alone or together with the visual stimulus. Results showed that spatially and temporally coincident visual stimuli improved sound localization accuracy. Consistent results have been observed also in infants (Morrongiello and Rocca, 1987). Moreover, these direct visual contributions to sound localization have been proved useful when training acoustic space perception (Rabini et al., 2019, Strelnikov et al., 2011, Valzolgher et al., 2020).

Considerably less research has instead investigated indirect visual contributions to sound localization. Yet, the idea that vision of the surrounding environment can provide useful spatial information for spatial hearing dates back to the 1970s, when it was termed 'visual facilitation' (Warren, 1970). Warren was among the first to report that sound localization accuracy can improve when listeners localize unseen auditory targets with their eyes open than closed. In eye-open conditions, speakers were hidden from the subject using a fabric screen, but the overall environment was clearly visible. When interpreting the advantage observed in eye-open conditions, Warren proposed that participants use visual cues from the environment to place the auditory stimulus into a visual spatial representation, instead of using only an auditory frame of reference. In line with this early observation, a decade later Shelton and Searle (1980) showed that when an speaker array is placed directly in front of the subjects, vision of the sources can facilitate accuracy compared to a blindfolded condition – even when the exact position of the target sound remains unknown. More recent works have replicated these pioneering studies, suggesting that seeing the environment can also enhance the precision of motor response and may thus affect participants' performance by facilitating their motor behavior. For instance, Redon and Hay (2005) showed that the presence of a visually structured background reduces pointing bias to visual targets. Interestingly, even the brief observation of the overall environment can improve spatial hearing (Tonelli et al., 2015).

Yet, in these classic studies as well as in more recent ones, it was difficult to disambiguate indirect contributions to sound localization that resulted from seeing the overall structure of the environment (e.g., a visible room), from the contributions resulting from seeing the possible space occupied by the sound sources (e.g., a panel hiding sound sources placed in front of the participant). In the first case, participants could code sound position with respect to existing references; in the second case, participants can develop some sort of visual prior about the position of the sounds in the environment (Parise et al., 2014). It is worth noting that in most studies exploring spatial hearing, the position of the sound sources is either directly visible or can be easily inferred. For instance, when all sources are hidden behind a curtain participants can nonetheless infer that sounds can originate from a restricted portion of the space (Pavani et al., 2017). In this case, although participants have no detailed prior about the spatial layout of the speakers, they have continuous visual priors about the hemispace (front or back), elevation and distance of the speaker array.

One way to disentangle between these indirect visual contributions to sound localization is to exploit virtual reality. In a recent study, Ahrens et al. (2019) asked participants to perform a sound localization task in different visual scenarios created using virtual reality technology. In some conditions, visual information about the room was entirely prevented. In other conditions, participants were allowed to see a virtual version of the real room, comprising or not the speakers around them. Using VR, these authors were able to control the effect of both having the *structure of the overall auditory environment* and knowing the *spatial likelihood* of the auditory targets. They found that the reference frame provided by the visual information of the room without loudspeaker was enough to decrease error both in the horizontal and vertical dimensions compared to a blindfolded condition. Moreover, vision of the speaker array provided a further benefit compared to receiving only visual information of the room. Along a similar line, Majdak et al. (2010) have manipulated the whole visual background during a sound localization task to study the impact of seeing a simple visual spatial frame (a grid) compared to a condition of total darkness. The results showed that even a simple grid improved precision in the horizontal plane. Furthermore, the visible grid reduced quadrant errors in the vertical plane, particularly in the front hemispace. In this work, sounds were always delivered using HRTF making more difficult to study the distance dimension. Likewise, the visible grid lacked binocular cues and did not provide information about distance. Finally, participants were forced into a static head listening posture. While this posture is most common in sound localization studies, it is a constraint that may limit the benefit from the visible visual environment on sounds localization. As proposed by some authors (Shelton and Searle, 1980), the spatial hearing facilitation that can result from seeing the overall environment, may reflect the active sensory-motor exploration of the auditory scene.

The present study aimed to test the contribution of a visual spatial frame to sound localization when participants are free to move their head without restriction. To this aim, we tested participants in two visual conditions: a uniform gray scene, in which no cues about the sound sources or the auditory environment were provided, and a simple visual spatial frame, in which a visible grid was the only visual information available about the environment. To present the different visual scenarios and to allow recording of spontaneous head and gaze movements, we took advantage of a new approach for the study of spatial hearing developed in our laboratories (Verdelet et al., 2020, Gaveau et al., 2020). Our approach is based on real sounds, virtual reality and real-time kinematic tracking and it allows: (1) accurate positioning of real sounds at pre-determined locations with respect to the head; (2) measuring the participant's hand responses in the three dimensions of space (i.e., in 3D); (3) control over the visual scenarios and (4) free and measurable head-movements and gaze during sounds playback. The latter aspect is particularly advantageous for the present study, as we hypothesized that head and eyes orienting to the sounds could provide implicit measures of performance to further investigate the effect of visual manipulation. A secondary purpose of the present work was also to validate the feasibility of our VR approach for the study of sound localization abilities in humans.

2. Methods

2.1. Participants

Thirty-six participants (age: M = 25.08, SD = 2.96, range [20-32], 13 males, 34 right-handed) were recruited to participate in the experiment at the University of Trento (Italy), at the Integrative, Multisensory, Perception, Action and Cognition Team (IMPACT) lab in Lyon and at the center de Recherche Cerveau et Cognition (CerCo) of Toulouse (France). All participants signed an informed consent before starting the experiment, which was conducted according to the criteria of the Declaration of Helsinki (1964, amended in 2013) and approved by the respective ethical committees. All had normal or corrected-to-normal vision and reported no movement deficit. Hearing thresholds were measured using an audiometer for all participants, testing different frequencies (250, 500, 1000, 2000, 4000 Hz), on the right and left ear separately. All participants had an average threshold below 20 dB HL.

2.2. Apparatus and stimuli

Virtual reality (VR) and kinematic tracking was implemented using 3 identical HTC Vive Systems, one for each testing site. This method

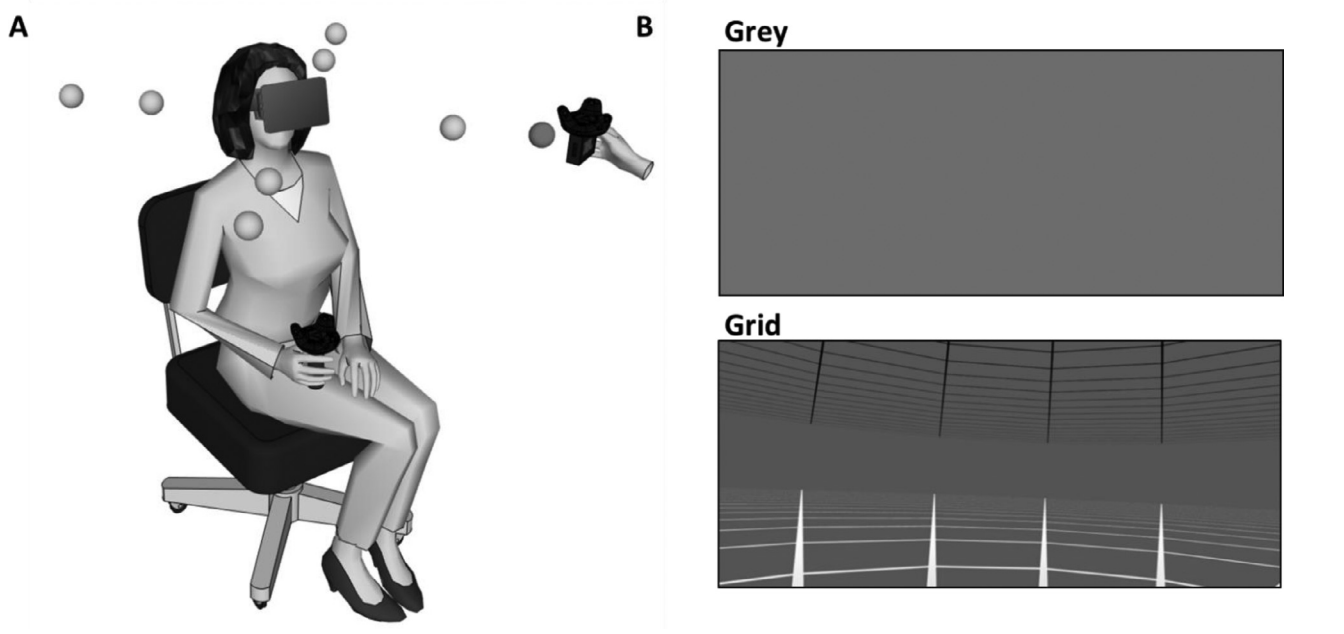

Fig. 1. Experimental procedure and setting. (A) Schematic representation of participant wearing the HMD and holding the tracker used for pointing responses; the eight spheres around the participant's head indicate the pre-determined speaker positions; the experimenter brought the tracked speaker (also shown in figure) at the pre-determined location identified in each trial. (B) Representation of the two 3D visual environments used in the study: gray and grid. A video showing dynamically the grid condition as seen from the participant's perspective is available at: https://youtu.be/89xPLzr3fyQ.

(European patent n°17723294.6–1115) has been developed in our laboratory (see also Gaveau et al., 2020 and Valzolgher et al. 2020). Each system (Fig. 1A) comprised one head-mounted display (HMD, resolution: 1080 × 1200 px, Field Of View (FOV): 110°, Refresh rate: 90 Hz), 1 controller (used by experimenter to calibrate head-position and to interrupt trial data registration), 2 trackers (one mounted on a short rod and held by participants to indicate the position of the sound and the other mounted above the speaker to track its position in real time) and 2 lighthouse base stations (scanning the position of the controller and trackers). Tracking precision and accuracy of the HTC Vive System is adequate for behavioral research purposes (Ahrens et al., 2019). The HMD was equipped with an SMI eye-tracking system (250 Hz). At all testing sites, stimuli were controlled and delivered using a LDLC ZALMAN PC (OS: Windows 10 (10.0.0) 64bit; Graphic card: NVIDIA GeForce GTX 1060 6GB; Processor: Intel Core i7–7700 K, Quad-Core 4.2 GHz/4.5 GHz Turbo - Cache 8 Mo - TDP 95 W) using Steam VR software and the development platform Unity3D (Unity Technologies, San Francisco, CA).

Participants were seated on a rotating armless chair with no chin rest, in the center of the room. The rooms had the following dimensions: Trento: 4 × 3 m; Lyon: 3 × 6 m; Toulouse: 3 × 5 m. All rooms were quiet, but none was specifically treated for being anechoic and sound-proof.

Real free field auditory stimuli were delivered by an unseen loudspeaker (JBL GO Portable, 68.3 × 82.7 × 30.8 mm, Output Power 3.0 W, Frequency response 180 Hz – 20 kHz, Signal-to-noise ration > 80 dB), whose position was continuously tracked in space. They were 3 s white noise bursts, amplitude-modulated at 4 Hz, and delivered at about 60 dB SPL, as measured from the participant's head (using Decibel Meter TES1350A placed at ears level).

This solution allowed us to track the position of the speaker, the hand of participant and the Head Mounted Display, via a dead reckoning process using gyroscope and accelerometer, plus a correction signal from the lighthouse system every 8.333 ms. Both tracking method allowed us to track this position with a frequency sample of 250 Hz. The software is designed to guide the experimenter to align the real loudspeaker (the sound source) with each of the 8 pre-determined position in the virtual environment in each trial.

2.3. Procedure

Before starting the experiment, participants were instructed about the task and informed about the use of the VR equipment. Specifically, participants were asked to listen carefully to each sound until it finished, and then to indicate its location in space using the tracker held in their right hand. They were informed that sounds could be delivered anywhere in the 3D around them, always within a reaching space. They were also informed that, during sound emission, they were not allowed to move their hand (which rested on their laps) nor their trunk (which was oriented straight ahead). After sound emission, they could freely rotate the chair to indicate the sound source. Note that head-movements remain unconstrained both during and after sound emission, allowing the possibility of spontaneous active listening behavior (e.g., orienting the head in the direction of the sound).

Participants performed sound localization under two visual conditions: a uniform gray scene (gray) and a more structured scene with spatial references (grid) (Fig. 1B and Video). The grid comprised two horizontally laid figures, drawn like spiderwebs of 50-meters radius, with 19 straight sides (angle around 20°) and 20 sub-figures plan separated by 2.5 m. The first horizontal web was placed at floor level (Y = 0 m) and the second was place at 10 m height (Y = 10 m). In creating the grid, we aimed to obtain a structured environment, with distance clue that conveyed the idea of a vast and open space (hence, the 50-meter radius of the spiderwebs). In addition, we aimed to avoid any vertical line that could be used by participants as visual anchor during sound perception and/or during the pointing response. The overall grids were centered on the participant's position. A video showing dynamically the grid condition as seen from the participant's perspective is available at: https://youtu.be/89xPLzr3fyQ (note that the video does not represent an experimental trial, but only depicts the environment visible to the participants).

After the participant sat on the chair in the experimental room and wore the HMD, the experiment began the head-center calibration which was performed by collecting the 3D position of the two ears using the controller. These head-center coordinates served as origin of the polar frame of reference that defined speaker, head and gaze positions

throughout the experiment. Then, the eye-tracker calibration was performed: participants were asked to follow a moving dot with their eyes. Both head and eye calibrations were repeated whenever the HMD was temporarily removed during the experiment (i.e., during pauses).

The loudspeaker position in 3D space was calculated for each trial, with reference to the center of the head. In this way, despite participants sat without any chin-rest, we could carefully control the position of each sound source with respect to the ears. Eight pre-predetermined positions were used throughout the experiment, resulting from the combination of 4 different azimuths (-45°, 45°, -135° or 135°), 2 different distances (35 cm or 55 cm) and a single elevation (0°, i.e., ear-level) (Fig. 1).

In each experimental trial, the experimenter moved the loudspeaker to the desired position in 3D space, following visual instruction generated in real-time by the computer. Instructions conveyed the predetermined azimuth and distance position for the speaker. These instructions were visible only to the experimenter, and they were delivered using the stimulus visualization monitor (IIYAMA PROLITE E2280HS 22″, resolution: 1980 × 1080, format 16:9) placed in the testing room. The monitor provided a bird-eye view of the experimental room and conveyed the pre-determined position of the sound source for that trial (as a red circle) and the actual real-time position of the speaker (as tracker picture). Using this visual instruction, the experimenter reached for computer-determined position rapidly, keeping the speaker approximately at ear-level. The computer considered the loudspeaker correctly positioned when it entered a sphere of 6 cm diameter centered on the pre-determined sound position.

The noises produced by the experimenter while placing the speaker were minimal. Nonetheless, participants were explicitly informed to pay attention to the target sounds, as any other sound in the room could be deceiving. To prove this point at the beginning of the testing session, the experimenter showed how she could stay to the right of participants while delivering the sound from their left by stretching the arm and hand holding the speaker. This was a demonstration that occasional noise cues about the movement of the experimenter could provide misleading information and thus discourage participants to rely upon this information. Most importantly, pilot work in our laboratory showed that participants (N=6) cannot reliably point to the speaker when the same placement procedure is used but no target sound is delivered. In this scenario the three-dimensional vector distance between the speaker and the response is on average 73.3 cm (SD = 31.7 cm).

The computer delivered the target sound only when three concurrent criteria were met: (1) the loudspeaker was in the 3D position predetermined for the trial; (2) the participant's head was facing straight ahead; (3) the participant's eyes were directed straight ahead. Participants complied with criterion 2 (head pointing straight ahead) and criterion 3 (eyes gazing straight ahead) by taking advantage of visual stimuli displayed in the HMD. At the beginning of each trial two crosses were presented to the participant: a white cross in the background indicated the desired position of the head and eyes; and a thin white cross, indicating the actual head-position of the participant. Participants were instructed to move their head to align the two crosses. When the alignment was achieved the thin cross turned blue. Likewise, participants were instructed to stare at the cross center, a feedback of their gaze location was given by a blue circle. Once the three criteria were achieved simultaneously (speaker position; head position and eye gaze), all visual feedback disappeared, and the sound was delivered. Participants were instructed to respond only after the end of the sound, bringing the tracker to the perceived location of the sound and holding it still a few seconds. The experimenter terminated the registration of the tracking by pressing a button on the controller. No feedback on performance was provided (similar procedure was used also in Gaveau's study (Gaveau et al., 2020), see also (Valzolgher et al., 2020).

The experimental session was organized in 4 successive blocks, with a pause between each block. Visual conditions (gray or grid) alternated between blocks of trials. Half of the participants followed a grid-gray-gray-grid sequence, whereas the other half followed a gray-grid-grid-gray sequence. Each block comprised 40 trials (i.e., 5 trials for each of the 8 pre-determined positions), resulting in a total of 160 trials (i.e., 10 trials for each pre-determined position in each visual condition). The entire experimental protocol lasted approximately 45 min.

2.4. Analyses

The position of all tracked elements (loudspeaker, head center and direction, hand and eyes) was inspected manually for each trial. Loudspeaker position was calculated as the mean of x, y, z coordinates from the beginning of the sound to the end. Head and hand positions were analyzed using custom-made software for the kinematic analysis of movements, running on MATLAB R2019b.

To study head and gaze movements, we calculated the tangential velocity on the x, y, z axis (expressed in degrees of rotation) using two-points central difference derivate algorithm (Terry Bahill and McDonald, 1983) with 5 points for the half-window. To determine the sequence of head and hand movements, the beginning and the end of all movements were automatically detected using a velocity threshold procedure (10°/s for head and 400 mm/s for eyes). The results of this procedure were inspected off-line and corrected manually, if necessary. This procedure served to establish the spatio-temporal profile of head and hand behavior and extract relevant parameters for subsequent analyses (reaction times (RT) of the first head movement and first eyes movement). Head movements below 10° were discarded. Importantly, it also served to reject all trials in which participants did not comply with the instructions (i.e., they made anticipatory hand movement during sound playback) or because of artefacts or lack of data.

All statistical analyses and data visualizations were performed using R, R-studio environment and JASP 0.9.1.0.

3. Results

Our approach to sound localization allowed to decompose localization errors in the three dimensions of space: azimuth, elevation and distance. We started by studying participants' ability to discriminate sound source location in the uniform gray condition – which we considered as baseline – and then we focused on our key experimental question, concerning the effects of seeing a visual grid. This two-step approach was motivated also by our interest to validate our VR approach to spatial hearing which, for several aspects, is novel with respect to other previous methods. Unlike classic works using real sounds, our approach allows free but carefully controlled positioning of the speaker in head-centered coordinates. Moreover, unlike classic auditory virtual reality studies, it uses actual sounds delivered in space and aligned with the visual reality scenario.

3.1. Hand pointing responses: the uniform gray condition

Sound localization responses for each participant are shown in Fig. 2A and B in bird-eye and lateral views, respectively, as a function of sound position. Participants clearly discriminated stimulus side (overall left/right discrimination errors = 1.0%, SD = 0.7), and made very few front/back confusions (overall errors = 1.4%, SD = 0.6). In addition, as shown in Fig. 2C, responses in distance were clearly segregated for near and far targets (near: M = 42.7 cm, SD = 10.8 cm; far: M = 58.6 cm, SD = 11.7 cm; t (35) = 16.15, $p < 0.001$).

For each dimension, we studied absolute and signed errors as a function of target position, using an Analysis of Variance (ANOVA) with SIDE (left or right), ANTERO-POSTERIOR SECTOR (front or back) and DISTANCE (near, 35 cm or far, 55 cm) as within-participants factors.

3.1.1. Azimuth

The overall absolute error in azimuth was 11.5° (SD = 6.5°). The analysis on absolute error in azimuth revealed a main effect of ANTERO-POSTERIOR SECTOR, F (1,35) = 6.00, $p = 0.02$, $\eta^2 = 0.07$, caused by larger

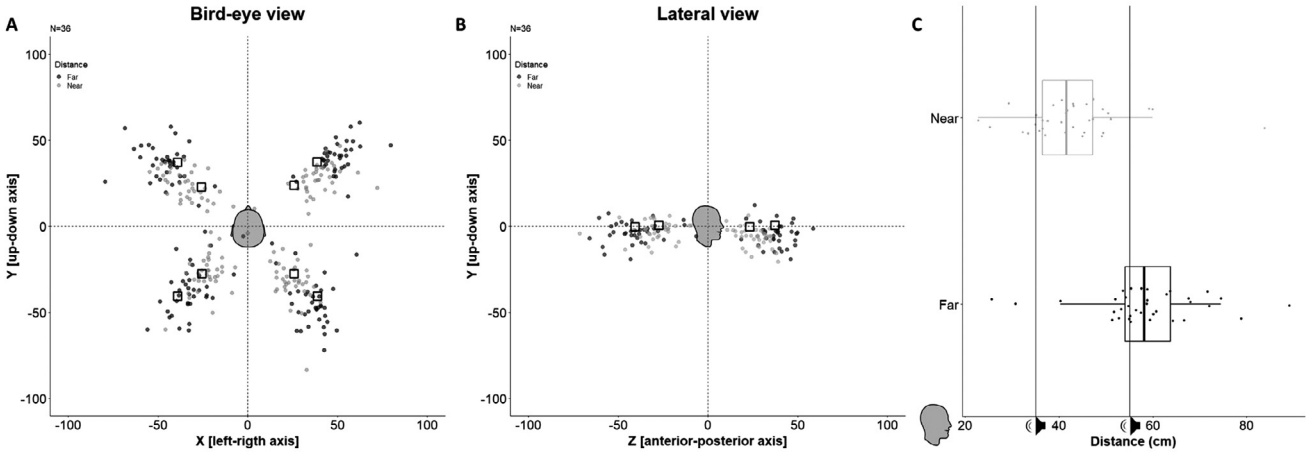

Fig. 2. Sound localization performance. (A) Bird-eye view of all target positions (squares with black border) and hand-pointing responses (smaller circles) in all trials and participants. (B) Lateral view of all target positions and responses. Responses for each participant are averaged across side (left or right) and distance (near, 35 cm and far, 55 cm). Responses are color-coded as a function of target distance (far is dark gray and near is light gray). (C) Distance of participants' hand-pointing as a function of target position (Near or Far). Vertical lines represent the real position of the targets (Near: 35 cm; Far = 55 cm).

errors in back space (M = 13.5°, SD = 9.1°) compared to front space (M = 9.4°, SD = 7.2°). The main effect of DISTANCE also reached significance, F (1,35) = 13.10, $p < 0.001$, $\eta^2 = 0.02$. Errors were larger for near (35 cm) (M = 12.6°, SD = 7.8°) compared to far (55 cm) targets (M = 10.4°, SD = 5.5°). No other main effect or interaction reached significance (all Fs < 2.51, all $ps > 0.12$).

A similar analysis on signed error (speaker position minus participant's tracker position, positive values indicate a rightward bias) revealed only a main effect of SIDE, F (1,35) = 26.39, $p < 0.001$, $\eta^2 = 0.22$. This main effect reflects the fact that participants' pointing responses were more eccentric than actual sound azimuth (left: M=-5.2°, DS=8.2°; right: M=5.3°, DS=6.4°) as shown in Fig. 2A. Thus, a positive (rightward) bias emerged for right sounds, where a negative (leftward) bias emerged for left sounds.

3.1.2. Elevation

The overall absolute error in elevation was 13.5 (SD = 5.4°). The analysis on absolute error in elevation revealed a main effect of ANTERO-POSTERIOR SECTOR, F (1,35) = 4.17, $p = 0.05$, $\eta^2 = 0.04$, caused by larger errors in front space (M = 14.9°, SD = 6.8°) compared to back space (M = 12.1°, SD = 6.7°). In addition, there was a main effect of DISTANCE, F (1,35) = 22.49, $p < 0.001$, $\eta^2 = 0.05$. Elevation errors were larger for near (M = 15.0°, SD = 5.6°) compared to far sounds (M = 12.0°, SD = 5.8°). No other main effect or interaction reached significance (all Fs < 3.27, all $ps > 0.08$).

A similar analysis on signed errors (positive values indicate an upward bias) revealed a main effect of ANTERO-POSTERIOR SECTOR, F (1,35) = 25.15, $p < 0.01$, $\eta^2 = 0.31$. Participants positioned the hand tracker below actual speaker location in front space (i.e., a downward bias; M = -9.2, SD = 10.5), but pointed above actual speaker location in back space (i.e., upward bias: M = 4.2, SD = 8.7). These biases were more pronounced for near than far targets, resulting in an interaction between ANTERO-POSTERIOR SECTOR and DISTANCE, F (1,35) = 9.37, $p = 0.004$, $\eta^2 = 0.01$.

3.1.3. Distance

The overall absolute error in distance was 12.8 cm (SD = 6.7 cm). The analysis on absolute error in distance revealed a main effect of SIDE, F (1,35) = 4.63, $p < 0.04$, $\eta^2 = 0.02$, caused by larger errors for right (M = 13.6 cm, SD = 7.5 cm) than left targets (M = 12.1 cm, SD = 6.6 cm). In addition, all 2-way interactions reached significance: SIDE and DISTANCE, F (1,35) = 13.43, $p < 0.001$, $\eta^2 = 0.01$; SIDE and ANTERO-POSTERIOR SECTOR, F (1,35) = 8.06, $p < 0.007$, $\eta^2 = 0.01$; ANTERO-POSTERIOR SECTOR and DISTANCE, F (1,35) = 15.69, $p < 0.001$, $\eta^2 = 0.02$. Overall these interactions capture the larger error in distance estimation that occurred for near targets in front right space (M = 16.6, SD = 10.2).

A similar analysis on signed errors in distance (positive values indicate overestimation of target distance) revealed the main effects of SIDE (left or right), ANTERO-POSTERIOR SECTOR and DISTANCE all (Fs > 9.87, all ps < 0.003). Overall, sound distance was overestimated (M = 5.8 cm, SD = 10.7 cm), but more for right (M = 9.1 cm, SD = 10.5 cm) compared to left sounds (M = 2.5 cm, SD = 11.5 cm). Overestimation was also larger for front (M = 11.0 cm, SD = 10.7 cm) compared to back sounds (M = 0.6 cm, SD = 12.7 cm), and for near (M = 7.2 cm, SD = 10.8 cm) compared to far sounds (M = 4.4 cm, SD = 11.3 cm). Several higher-order interactions also reached significance: SIDE and ANTERO-POSTERIOR SECTOR, F (1,35) = 5.65, p < 0.02, $\eta^2 = 0.003$, SIDE and DISTANCE, F(1,35) = 5.46, $p < 0.03$, $\eta^2 = 0.002$, and between SIDE, ANTERO-POSTERIOR SECTOR and DISTANCE, F (1,35) = 22.57, $p < 0.001$, $\eta^2 = 0.005$. This 3-way interaction did not reveal any difference between near and far signed errors targets when placed in the back-right sector.[1]

3.2. Hand pointing responses: the effects of adding a visual frame

Having established performance in the three dimensions in the gray visual condition, we turned to examine to what extent adding a visual frame affected sound localization performance. To this aim, we entered absolute and signed errors into ANOVAs with VISUAL CONDITION (gray, grid) and ANTERO-POSTERIOR SECTOR as within-participant variables, separately for each spatial dimension. We focused on the antero-posterior position of sounds because we predicted that the effect of the

[1] We interpreted these differences observed for near sounds in right space as the consequence of bio-mechanical constraints related to the fact that all participants were asked to point to the sounds using their right hand. To assess this hypothesis empirically, we analyzed the data of a pilot experiment in which 12 participants performed the same task reported in the present manuscript but using their preferred hand on a trial-by-trial basis. An ANOVA on absolute errors in distance on this dataset, using again SIDE, ANTERO-POSTERIOR SECTOR and DISTANCE as factors, yielded no significant main effect or interaction (all Fs < 3.20, all ps > 0.10). On signed errors in distance, we found only a significative interaction between ANTERO-POSTERIOR SECTOR and DISTANCE, F (1,11) = 7.22, $p = 0.02$, $\eta^2 = 0.03$. This provides initial evidence that at least part of the specificities observed for right-sided sounds could depend upon the imposed use of the right hand for the response.

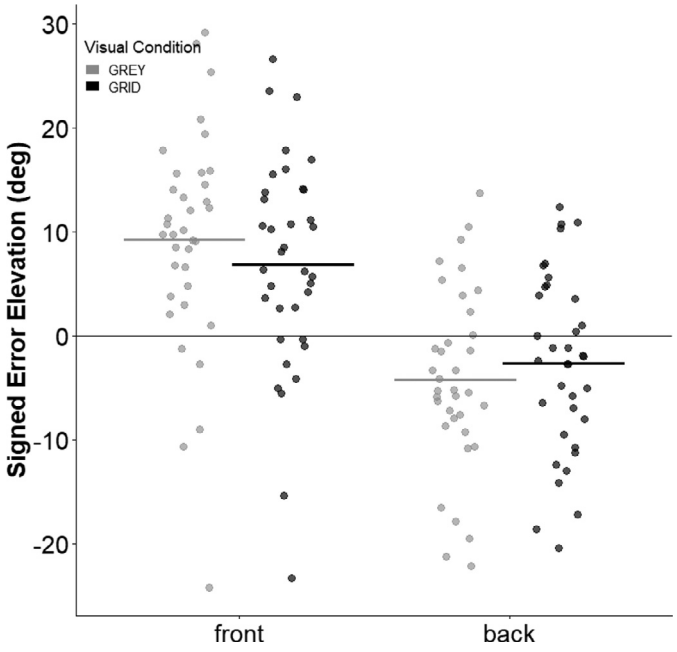

Fig. 3. Signer Error in Elevation (deg) as a function of Visual condition (gray, grid) and antero-posterior sector (front, back). In both graphs horizontal bars represent the mean of each condition, while points show participants value. *=p<0.05;**=p<0.01;***=p<0.001.

visual manipulation could be maximal for sounds were delivered in the front space rather than back space.

The analyses on absolute and signed errors in azimuth, elevation and distance, did not reveal any main effect or interaction involving VISUAL CONDITION (all *Fs* < 1.97, all *ps* > 0.17). One notable exception was signed error in elevation. For this measure, we found a significant interaction between ANTERO-POSTERIOR SECTOR and VISUAL CONDITION, $F(35) = 12.72$, $p = 0.001$, $\eta^2 = 0.27$. When sounds were delivered in the front space, participants judged their position lower than their actual location. This bias was smaller in the grid condition (M=-6.8°, DS = 10.3°) compared to the gray condition (M=-9.2°; DS = 10.5°; simple main effect: $F = 12.12$, $p = 0.001$). When sounds were delivered in the back space, participants judged their position as higher than their actual location. Again, this bias smaller in the grid condition (M=2.7°, DS = 8.6°) compared to the gray condition (M=4.2°; DS = 8.7°; simple main effect: $F = 5.84$, $p = 0.02$) (Fig. 3 and Table 1).

3.3. Head rotation

Head-movements occurred in 68.3% of trials on average (SD = 38.3%; median = 89.9, range = 0–100%), for targets in front and back space (71.3% and 65.3% of trials, respectively). Fig. 4A shows the substantial variability of this spontaneous behavior in both visual conditions. In front space, the first head movement for sounds at 45° to the right was of 40.1° (SD = 7.2°), whereas for sounds at 45° to the left it was 39.2° (SD =7.3°). In back space, the first head rotation for sounds at 135° to the right was of 67.9° degrees (SD = 21.0°), whereas for sounds at 135° to the left it was 68.7° (SD = 22.5°) (see Fig. 4B).

To analyze head movements rotation around the vertical axis, we considered two variables: 1) rotation amplitude of the first head movement (deg) and 2) time to first head-movement (the time between the beginning of the sound emission and the beginning of the first head movement in millisecond). Trial without head movements were excluded from these analyses. This resulted in the exclusion of 6 participants which did not move at all and 5 participants for whom less than 12 trials per condition remained. In addition, we focused only on those trials in which participants rotated towards the correct sound side. This occurred on 97.5% (SD = 2.0) of trials on average, indicating again that participants easily disambiguate sound side.

To test the effect of the presence of the grid on these two variables (head-rotation latency and amplitude), we entered each measure separately in an ANOVA with VISUAL CONDITION and ANTERO-POSTERIOR SECTOR as within variables. ANOVA on head-rotation amplitude revealed no main effect or interaction (all *Fs* < 1.26, all *ps* > 0.27). The analysis on time to first head-movement revealed that participants started to rotate their head 978 milliseconds (SD = 155) after sound emission on average. In addition, a significant main effect of ANTERO-POSTERIOR SECTOR emerged, $F(1,24) = 7.90$, $p = 0.02$, $\eta^2 = 0.25$. When sounds were emitted in front space, participants responded faster (M= 957 ms, SD = 154, ms) compared to when they were emitted from the back (M= 999 ms, SD = 156 ms) (Fig. 5A).

3.4. Gaze direction

Similar to head rotation analyses, we focused on trials in which participants gazed towards the same side of the sound (correct trials). This resulted in the exclusion of 10 participants for whom less than 12 trials per condition remained. On average participants gazed towards the same side of the target on 97.2% (SD = 1.7) of trials. Fig. 4C shows gaze direction as a function of target position. In front space, first gaze was directed to 42.1° (SD = 18.9°) for sounds 45° right, and to 44.6° (SD = 20.7°) for sounds 45° left. In back space, first gaze was directed 67.2° (SD = 27.3°) for sounds 135° right, and to 70.5° (SD = 25.2°) for sounds 135° left.

To analyze gaze movements, we considered two variables: (1) direction of the first gaze movement (deg) and (2) time to first gaze-movement (the time between the beginning of the sound emission and the beginning of the first gaze movement in millisecond).

To test the effect of our visual manipulation we entered these two dependent variables in separate ANOVAs, with VISUAL CONDITION and ANTERO-POSTERIOR SECTOR as within variables. The ANOVA on direction of the first gaze movement revealed a significant two-way interaction, $F(1,25) = 5.70$, $p = 0.03$, $\eta^2 = 0.19$, caused by a less symmetrical gaze directions in response to sounds in back space (M = -3.4°, DS = 9.0°), specifically in the grid condition (simple main effect: $F = 5.01$, $p = 0.03$).

A similar ANOVA on time to first gaze-movement revealed that participants started to gaze sounds 698 milliseconds (SD = 164) after sound emission. In addition, we found a significant main effect of ANTERO-POSTERIOR SECTOR, $F(1,25) = 19.54$, $p < 0.001$, $\eta^2 = 0.44$. Unsurprisingly, participants oriented their gaze faster to front (M= 651 ms, SD = 153 ms) compared back sounds (M= 744 ms, SD = 174 ms) (Fig. 5B). More interestingly, we also found a significant main effect of VISUAL CONDITION, $F(1,25) = 12.85$, $p = 0.001$, $\eta^2 = 0.34$. When the visual scenario was the grid, participants latency was shorter (M= 666 ms, SD = 146 ms) compared to when it was gray (M= 728 ms, SD = 271 ms). To test specifically whether VISUAL CONDITION had a greater impact in front than back space we run simple main effects. These revealed that the effect of VISUAL CONDITION was more substantial in the front ($F = 19.46$, $p < 0.001$) than back space ($F = 3.82$, $p = 0.06$; note that the 2-way interaction between VISUAL CONDITION and ANTERO-POSTERIOR SECTOR also approached significance: $F(1,25) = 6.46$, $p = 0.075$, $\eta^2 = 0.12$) (Fig. 5B).

4. Discussion

The present study examined the effect of seeing a simple visual spatial frame on sound localization, in a context in which participants were free to move their head and eyes while listening to sounds. To this aim, we leveraged a virtual reality (VR) approach which allows accurate control over the visual scenarios, accurate playback of free field sounds in three-dimensional space, and free and measurable movements of head and gaze during sounds playback.

Table 1
Absolute error and Signed error of 3 dimensions of the space as a function of target position for grid and gray visual conditions. Standard deviation (SD) between brackets.

			Azimuth (deg)		Elevation (deg)		Distance (cm)	
			Absolute Error	Signed Error	Absolute Error	Signed Error	Absolute Error	Signed Error
GRAY								
	NEAR							
	FRONT							
		Left	10.3 (7.8)	−5.8 (9.2)	16.0 (8.1)	−12.0 (10.9)	12.8 (9.0)	9.8 (12.1)
		Right	11.5 (13.1)	4.9 (13.5)	17.7 (9.5)	−10.0 (15.4)	16.6 (10.2)	15.8 (10.9)
	BACK							
		Left	15.8 (13.2)	−5.7 (16.0)	14.0 (8.8)	5.1 (12.3)	10.4 (7.6)	−1.0 (12.1)
		Right	12.8 (9.6)	6.6 (11.1)	12.2 (7.6)	4.7 (10.5)	11.8 (9.3)	4.4 (13.3)
	FAR FRONT							
		Left	8.5 (6.1)	−4.5 (7.5)	12.7 (6.5)	−8.0 (10.7)	11.9 (9.3)	6.7 (12.9)
		Right	7.5 (6.2)	4.2 (5.3)	13.2 (8.1)	−6.7 (12.4)	14.0 (8.3)	11.8 (10.2)
	BACK							
		Left	13.7 (12.9)	−4.7 (11.4)	11.5 (10.1)	3.2 (7.4)	13.2 (8.8)	−5.2 (14.6)
		Right	11.7 (7.2)	5.6 (10.1)	10.6 (7.2)	3.8 (10.3)	12.1 (8.5)	4.2 (13.5)
GRID								
	NEAR							
	FRONT							
		Left	11.2 (8.8)	−4.9 (9.4)	15.4 (8.3)	−8.7 (12.6)	12.9 (8.1)	10.8 (10.6)
		Right	9.3 (7.0)	6.4 (6.5)	15.7 (8.9)	−8.8 (14.0)	17.5 (10.1)	16.3 (11.7)
	BACK							
		Left	14.8 (11.2)	−9.0 (14.2)	13.8 (7.9)	1.7 (12.8)	10.3 (7.7)	−1.2 (12.4)
		Right	12.4 (7.6)	7.3 (10.2)	12.2 (7.5)	3.8 (10.7)	12.6 (8.1)	5.1 (12.2)
	FAR FRONT							
		Left	8.9 (6.4)	−4.9 (8.7)	11.8 (6.4)	−5.3 (10.1)	11.4 (8.6)	5.7 (12.8)
		Right	7.4 (8.8)	3.7 (6.0)	12.6 (8.8)	−4.6 (11.2)	14.3 (9.1)	12.0 (11.7)
	BACK							
		Left	13.6 (9.5)	−6.3 (12.1)	11.6 (10.7)	2.6 (12.0)	13.9 (9.1)	−6.6 (12.9)
		Right	11.2 (5.6)	7.7 (7.4)	9.8 (7.3)	2.5 (8.8)	12.5 (10.3)	3.7 (15.3)

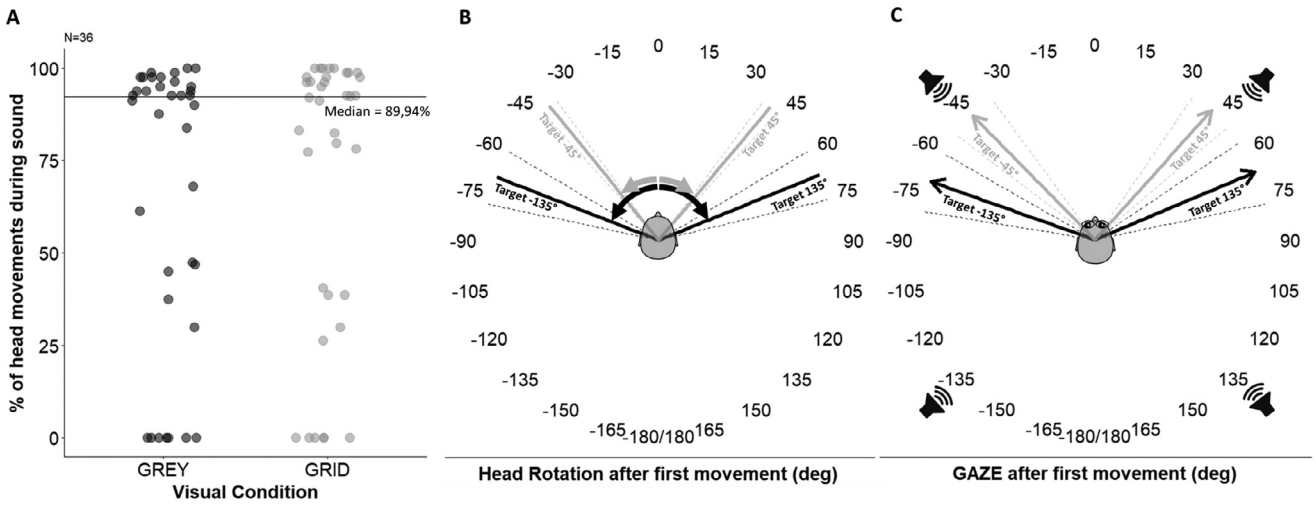

Fig. 4. (A) Boxplot of percentage head-movements: the percentage of trials in which head movements occurred for each participant. (B) Average of the rotation around vertical axis of the first head movement only when participants turned to the side of the target (same dataset of the analysis). (C) Average of the direction of the first gaze movement only when participants turned to the side of the target (same dataset of the analysis). Dashed lines represent confidence intervals.

When no visual spatial frame was available (gray condition), sound localization in azimuth was better in front than back space. When placing the tracker in back space participants had to turn around and modify their posture. This adjustment entails re-coding of sound position from egocentric to world-centered coordinates, leading to more uncertain estimates of sound position (Kopinska and Harris, 2003, Pavani et al., 2008). In addition, during the gray condition the participants' responses were characterized by a bias to point to positions more eccentric than actual speaker locations. Errors in azimuth were higher for far rather than near targets. Although elevation was not varied, errors in the vertical dimension were larger for near than far sounds, and frontal sounds were perceived as lower than their actual position. Sound distance was overestimated, particularly for front and right sounds. Concerning distance, it is important to notice that our approach led us to measure sounds localization abilities also considering this dimension.

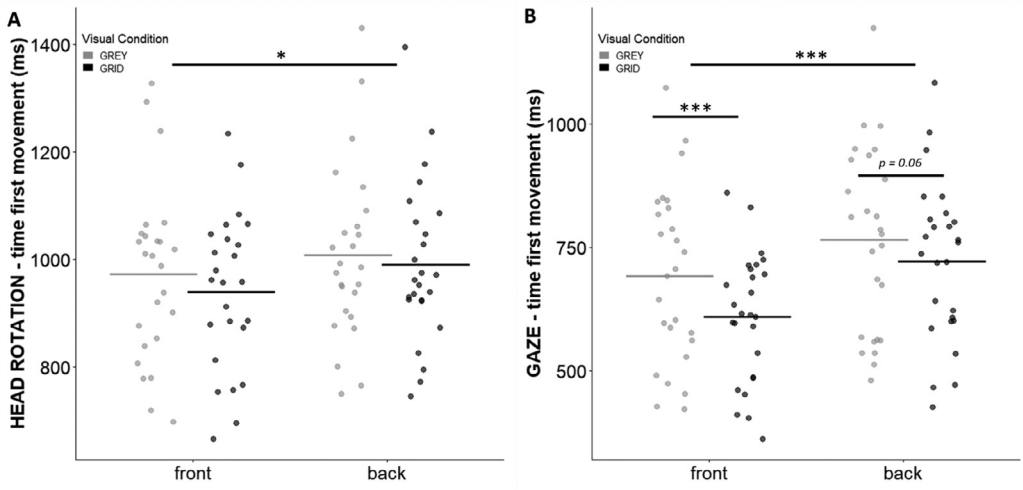

Fig. 5. T, the time between the beginning of the sound stimulation and the beginning of the first head or eyes movement. (A) Boxplot of Head Rotation time firs movement as a function of target position (front, back) color coded as a function of Visual Condition (gray, grid). (B) Boxplot of GAZE time first movement as a function of target position (front, back) color coded as a function of Visual Condition (gray, grid).

Adding a visual spatial frame (grid condition) did not change the overall pattern of hand pointing responses in azimuth or distance. However, it resulted in improved accuracy in elevation and it affected participant's' gaze behavior, producing faster orienting responses to the sounds.

4.1. Studying 'visual facilitation' with virtual reality

The question on the role played by visuo-spatial references on sound localization dates back to the 1970s (Warren, 1970), when it was termed 'visual facilitation'. Yet, until the advent of VR methods, visual facilitation could only be addressed with somewhat crude manipulations (e.g., blindfolding participants). Most importantly, when visual information about the environment was provided (i.e., eyes-open conditions), it was a methodological challenge to disentangle the contributions related to seeing the overall structure of the environment, from the contribution related to seeing the potential sources of sound. Using a VR approach, we investigated the role of visually structured information on sound localization, in the absence of visual priors about the speakers' position. While immersed in the VR scene, participants were only informed that target sounds would be delivered within reaching distance, but had no further information on their positions — i.e., they expected sounds to appear all around the body but inside an estimated range of distance (Gaveau et al., 2020).

To the best of our knowledge, only two previous works have addressed a similar question using VR techniques (Ahrens et al., 2019, Majdak et al., 2010). Majdak et al. (2010) tested the effect of seeing a simple visual environment on sound localization. Subjects were immersed in a virtual environment which comprised a sphere (diameter 5 m), marked with grid lines every 5° horizontally and every 11.25° vertically. Furthermore, participants required to judge elevation and azimuth of the sound taking advantage of grid lines. They reported that participants' hand pointing to sounds was enhanced by the visual grid in both horizontal and vertical dimensions, compared to a condition of total darkness. Our results are consistent with these findings. However, in the present study we observed an advantage only in terms of elevation, whereas sound localization in azimuth remained unchanged in the presence of the visual grid. This different result could reflect methodological discrepancies between the two studies. First, our visual grid was intentionally conceived to avoid any vertical line that could serve as anchor for sound localization. Second, our virtual environment conveyed also a sense of distance, which could have introduced additional uncertainty in the interpretation of the auditory cues. Third, we used real sounds instead of virtual ones, possibly leading to more precise sound localization overall.

Nonetheless, both studies converge in stressing the importance of having a visual frame when localizing sound. In particular, the grid condition may have provided a detailed visual map for positioning sound sources.

Ahrens and colleagues also took advantage of VR (Ahrens et al., 2019), allowing participants to see the room dimensions as well as their hand-position, whereas the speakers' array was not visible. Using sounds beyond reaching distance (2.4 m from the head), they found that visual information decreased errors both in the horizontal and vertical dimensions, as compared to a blindfolded condition. In Ahrens et al.'s study the VR reconstructed room was a careful visual replica of the in which the experiment was conducted, with the foam wedges of the anechoic chamber providing extremely rich vertical and horizontal visual references all around the participant. As discussed above for the study by Majdak and colleagues, it may be that these substantial visual cues played a role in the improvements observed in the horizontal dimension, serving as references or place-holders for sound localization when visible. In our more minimalist visual scenario we intentionally avoided all vertical visual references.

One difference with respect to both these previous works is that we presented sounds within reaching distance. In Majdak et al., sounds were delivered through headphones and although they were likely perceived externalized, it is difficult to establish at which distance they were perceived. In Ahrens et al., sounds were instead delivered away from the body, at 2.4 m. Our rationale in presenting sounds within reaching distance was to allow participants to respond using the hand-held tracker and measure their accuracy in distance estimation. However, this choice limited our studied space to the near-field and potentially influenced the interactions with the visual environment we created. Future works could examine to what extent the impact of visual environment on sound localization could emerge differently for sounds at different distances from the body, or — in case of enclosed spaces like rooms — at different relative distances with respect to the visible surfaces.

Another important difference is that our participants were free to move their heads while listening to sounds. This spontaneous orienting behavior also involved the eyes, and likely made azimuth localization easier for our participants. In turn, this could have reduced the possibility of observing visual condition effects in azimuth. However, this gaze (i.e., head and eyes) orienting behavior revealed interesting findings.

We found that correct gaze orienting responses in the direction of the sound started earlier when exposed to a visual grid compared to the gray condition. A similar trend was observed also for head rotation. We believe that the interest of analyzing this measure is related to the fact that, unlike hand pointing, gaze orienting is a more implicit measure of sound localization performance — particularly when the first gaze to sound response is considered. Measuring participants' gaze allowed us to capture an early and implicit aspect of acoustic space perception, which differs from the explicit head pointing method that has been used in sound localization tasks as an alternative to hand pointing. Taking this perspective, our results suggest that even a minimal visual spatial frame can speed up the right-left implicit disambiguation of sound position.

These effects of the visual features of the environment on sound localization are complementary with the line of research that examined the effect of the *acoustic* feature of the environment on sound localization and visual scene perception. For instance, Gil-Carvajal et al. (2016) have examined the effect of a mismatch between playback and recording room on perceived distance and azimuth direction of sounds. They found that sound distance ratings decreased when measured in an environment that was more reverberant than the original recording room, whereas azimuth direction remained unaffected. Most interestingly, they also observed that changes in the visual attributes of the room were ineffective and concluded that visual congruency is less crucial that the correspondence between the acoustical features of environment and the stimuli.

Other works examined the interplay between auditory and visual features of the environment. Etchemendy et al. (2017) have found that auditory reverberation cues impact on the perception of visible room size, revealing that the auditory environmental context can modulate visual distance perception. At the same time, Schutte et al. (2019) have shown that visual room impression does not affect people's abilities to estimate rooms' reverberations. The latter evidence in particular is directly relevant to the present work, as it suggests that the effect of the visual grid proposed in our experiment may not have influenced the extraction of the acoustic features from the environment.

In the present study conducted in three different experimental sites, we did not measure the rooms' reverberation limiting the possibility to deepen the acoustic influences on both sound localization performance and visual manipulation effects. Further work should aim to combine the contribution of visual and auditory features of the environment with the active sound localization approach introduced by the present work, to examine to what extent the relative contribution and potential interactions between these multisensory contextual contributions to spatial hearing. It would also be important to assess sound localization beyond reaching (as here) to address more directly whether sounds that perceived further away from the head (e.g., at 2.4 m as in Ahrens et al., 2019 or beyond) could interact more with the wider spaces of rooms or open VR environments like the ones simulated in our grid condition.

5. Conclusion

In this study, we documented that providing visual frame that is totally uninformative about sound source position in space helps sound source localization in active listening conditions. These findings contribute to emphasize the indirect but positive contribution of minimally structured vision to sound localization and further promote the idea that sound localization ability should be conceived as a multisensory process. Our findings also underlie the importance of allowing and measuring spontaneous head and gaze movements. The here adopted VR technology allowed us to go beyond traditional approaches in the study of spatial hearing by allowing participants to move their eyes and head during and after sound presentation, while retaining full control over sound placement and recording of our dependent variables using ecologically valid contexts, which is crucial in studying hearing experience as suggesting by recent studies in the field (Hadley et al., 2019).

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgements

We thank all the participants and the colleagues who helped during research. We thank Giordana Torresani for graphic solutions (Fig. 1A). **Funding**: C.V. was supported by a grant of the Università Italo-Francese (UIF)/Université Franco-Italienne (UFI) and the Ermenegildo Zegna Founder's Scholarship. F.P., C.V. and AF were supported by a grant of the Agence Nationale de la Recherche (ANR-16-CE17-0016, VIRTUAL-HEARING3D, France) and by a prize of the Fondation Medisite (France). The study was supported by the IHU CeSaMe ANR-10-IBHU-0003 and it was performed within the framework of the LABEX CORTEX (ANR-11-LABX-0042) of Université de Lyon. We thank JL Borach, S Alouche, S Terrones for administrative support, and Eric Koun for informatics support.

References

Ahrens, A., Lund, K.D., Marschall, M., Dau, T., 2019. Sound source localization with varying amount of visual information in virtual reality Axel. PLoS ONE 14 (3) Mar.
Alais, D., Burr, D., 2004. The ventriloquist effect results from near-optimal bimodal integration. Curr. Biol. 14 (3), 257–262 Feb.
Bolognini, N., Leor, F., Passamonti, C., Stein, B.E., Làdavas, E., 2007. Multisensory-mediated auditory localization. Perception 36 (10), 1477–1485 Oct.
Etchemendy, P.E., et al., 2017. Auditory environmental context affects visual distance perception. Sci. Rep. 7 (1), 1–10 Dec.
Gaveau, V., et al., 2020. SPHERE: a novel approach to 3D and active sound localization. bioRxiv p. 2020.03.19.998906, Mar.
Gil-Carvajal, J.C., Cubick, J., Santurette, S., Dau, T., 2016. Spatial hearing with incongruent visual or auditory room cues. Sci. Rep. 6 (1), 1–10 Nov.
Hadley, L.V., Brimijoin, W.O., Whitmer, W.M., 2019. Speech, movement, and gaze behaviours during dyadic conversation in noise. Sci. Rep. 9 (1), 1–8 Dec.
Heffner, R.S., Heffner, H.E., 1992. Visual factors in sound localization in mammals. J. Comp. Neurol. 317 (3), 219–232 Mar.
Heffner, H.E., Heffner, R.S., 2014. The Behavioral Study of Mammalian Hearing. Springer, New York, NY, pp. 269–285.
Hofman, P.M., Van Riswick, J.G.A., Van Opstal, A.J., 1998. Relearning sound localization with new ears. Nat. Neurosci. 1 (5), 417–421.
Knudsen, E.I., Knudsen, P.F., 1985. Vision guides the adjustment of auditory localization in young barn owls. Science (80-.). 230 (4725), 545–548.
Kopinska, A., Harris, L.R., 2003. Spatial representation in body coordinates: evidence from errors in remembering positions of visual and auditory targets after active eye, head, and body movements. Can. J. Exp. Psychol. 57 (1), 23–37.
Majdak, P., Goupell, M.J., Laback, B., 2010. 3-D localization of virtual sound sources: Effects of visual environment, pointing method, and training. Attent. Percept. Psychophys. 72 (2), 454–469 Feb.
Morrongiello, B.A., Rocca, P.T., 1987. Infants' localization of sounds in the horizontal plane: effects of auditory and visual cues. Child Dev. 58 (4), 918 Aug.
Parise, C.V., Knorre, K., Ernst, M.O., 2014. Natural auditory scene statistics shapes human spatial hearing. Proc. Natl. Acad. Sci. U. S. A. 111 (16), 6104–6108 Apr.
Pavani, F., Husain, M., Driver, J., 2008. Eye-movements intervening between two successive sounds disrupt comparisons of auditory location. Exp. Brain Res. 189 (4), 435–449 Aug.
Pavani, F., et al., 2017. Spatial and non-spatial multisensory cueing in unilateral cochlear implant users. Hear. Res. 344, 24–37.
Rabini, G., Altobelli, E., Pavani, F., 2019. Interactions between egocentric and allocentric spatial coding of sounds revealed by a multisensory learning paradigm. Sci. Rep. 9 (1) Dec.
Redon, C., Hay, L., 2005. Role of visual context and oculomotor conditions in pointing accuracy. Neuroreport 16 (18), 2065–2067.
Schutte, M., Ewert, S.D., Wiegrebe, L., 2019. The percept of reverberation is not affected by visual room impression in virtual environments. J. Acoust. Soc. Am. 145 (3), EL229–EL235.
Shelton, B.R., Searle, C.L., 1980. The influence of vision on the absolute identification of sound-source position. Percept. Psychophys. 28 (6), 589–596 Nov.
Strelnikov, K., Rosito, M., Barone, P., 2011. Effect of audiovisual training on monaural spatial hearing in horizontal plane. PLoS ONE 6 (3), e18344 Mar.
Terry Bahill, A., McDonald, J.D., 1983. Frequency limitations and optimal step size for the two-point central difference derivative algorithm with applications to human eye movement data. IEEE Trans. Biomed. Eng. BME-30 (3), 191–194.

Tonelli, A., Brayda, L., Gori, M., 2015. Task-dependent calibration of auditory spatial perception through environmental visual observation. Front. Syst. Neurosci. 9, 84 no. JuneJun.

Valzolgher, C., Campus, C., Rabini, G., Gori, M., Pavani, F., 2020a. Updating spatial hearing abilities through multisensory and motor cues. Cognition 204, 104409 Nov.

Valzolgher, C., Verdelet, G., Salemme, R., Lombardi, L., Gaveau, V., Farné, A., Pavani, F., 2020b. Reaching to sounds in virtual reality: A multisensory-motor approach to promote adaptation to altered auditory cues. Neuropsychologia, 107665 doi:10.1016/j.neuropsychologia.2020.107665.

Verdelet G. et al., "Assessing spatial and temporal reliability of the vive system as a tool for naturalistic behavioural research," 2020, pp. 1–8.

Warren, D.H., 1970. Intermodality interactions in spatial localization. Cogn. Psychol. 1 (2), 114–133 May.