PROCEEDINGS of the
23rd International Congress on Acoustics

9 to 13 September 2019 in Aachen, Germany

Reaction times in multisensory localization tasks

Yi ZHOU[1]; Colton CLAYTON[2]

[1] Arizona State University, USA

[2] Arizona State University, USA

ABSTRACT

The encoding of space between the visual and auditory systems does not always align. For foveal species like humans and monkeys, the visual field is restricted to frontal space, whereas the auditory field is panoramic, covering the entire surrounding space. Sensitivity to sounds coming from the rear brought by spatial hearing is critical for avoiding unseen dangers. Our recent study shows that, in humans, vision's influence on auditory perception can extend to unseen rear space (Montagne and Zhou, 2018). The present study further investigates the effects of visual stimulation on reaction times when listeners localize an auditory target presented from the front or rear. Experiments are designed to survey two types of reaction times simultaneously: (1) choice reaction time for pushing a button to indicate the perceived front or back location of a sound; (2) saccade reaction time by listeners shifting their gaze to indicate the lateral direction of a perceived sound source. Our results show domain-specific effects of visual capture on both types of reaction times, and support our previous findings that audio-visual interactions are not limited by the spatial rule of proximity.

Keywords: Multisensory, Reaction Time, Sound Source Localization

1. INTRODUCTION

Sensory experience builds upon a multisensory analysis of many cues that all describe properties of the same objects in an environment (1). This multisensory strategy helps build an internal map of the physical environment to enhance perceptual analysis (e.g., who is talking, where is (s)he?) and guide actions (e.g., escape or attack). Sound source localization is an excellent example of this strategy. In everyday interactions with the environment, our initial reaction to the sudden onset of an unexpected sound (e.g., a quickly-passing vehicle) is to estimate its source location to calibrate a reaction better. The speed and accuracy of this reaction depend on the coordination of auditory and visual spatial functions. Vision's effects on auditory localization are manifested in both the spatial and temporal response domains. When sound and light are presented simultaneously, but from different locations, (1) the direction of a light biases the perceived location of a sound source (2-7), i.e., the "Ventriloquist Effect", and (2) a flash of light delays auditory localization to a greater extent than sound delays visual localization (8, 9). In a simple detection task, visual percepts can overwhelm audition to such an extent that some subjects do not even notice the presence of a sound, i.e., the "Colavita Effect" (10).

While it is commonly accepted that multisensory integration becomes stronger when sensory inputs are spatially and temporally closer to each other, the space encoded by these two modalities does not always align. For foveal species like humans and monkeys, the visual field is restricted to the frontal space, whereas the panoramic auditory field covers the entire frontal and rear space. The rear sensitivity provided by spatial hearing is critical for avoiding unseen danger coming from behind. Rear space, however, has been largely overlooked in multisensory research. Our recent study shows that in humans, vision's influence on auditory perception can extend to unseen rear space and interacts closely with auditory front-back confusions (FBCs) in auditory localization (11). The FBC errors are responses to the approximately correct angular displacement (relative to the midline) but in the wrong front-back hemifield (12, 13). While interaural time and level differences (ITDs and ILDs) offer the primary information about the horizontal angle of a sound source (14), front-back errors are rooted in the fact that these binaural difference cues do not, on their own, correspond with only one sound source location, but a host of them along the "cones of confusion" (15). We found that frontal visual cues can reduce

[1] yizhou@asu.edu
[2] ctclayto@asu.edu

FBC errors for frontal auditory targets and increase FBC errors for rear auditory targets due to visual capture to the frontal field (11). What remains unknown, however, is whether vision also affects reaction time (RT) in panoramic auditory localization. This study addresses this question. The test involves a dual-response task, where a listener was asked to localize the lateral, left-right direction of a free-field sound using eye movement and immediately after, decide the front-back direction of the sound by pressing one of two buttons arranged vertically on a game console. The results show that change in RT is correlated with the perceived modality of stimuli. For listeners following the sound direction, a visual LED stimulus delayed their RT, whereas for listeners following the light direction, light speeded up their RT, relative to auditory-only responses. The changes in RT were more significant for rear sounds.

2. METHODS

2.1 Listeners

Eight participants completed all sessions (3 male, age range from 19-34 years old, median 21.5 yrs). All were reported right hand dominant. All participants had normal hearing sensitivity as verified by standard audiometric techniques under insert earphones and reported normal or corrected-to-normal vision. Participants provided written, informed consent, and received financial compensation for their participation. The experiment was conducted in accordance with procedures approved by the Arizona State University's *Institutional Review Board*.

2.2 Apparatus and Stimuli

The free-field sound localization task was conducted in a double-walled, sound-attenuated chamber (Acoustic Systems RE-243, 2.1 m x 2.1 m x 1.9 m) lined with 3" acoustic foam. The participant was seated in the center of the sound chamber with his/her head stabilized using a high-precision head positioner (HeadLock^{TM}, Arrington research). Sound stimuli were presented from four loudspeakers (full-range monitor, Adams F5, positioned at +/− 35^o and +/− 145^o at a distance of 1.1 m) hidden behind a black, acoustically transparent curtain. Light stimuli were delivered via a high-resolution LED bar attached to the acoustic curtain spanning +60^o to −60^o. For this experiment, only three active LED locations were used for testing (0^o, +/− 8^o). All sound and light stimuli originated from 0^o elevation relative to eye level.

Custom-designed software written in MATLAB generated and controlled the auditory stimuli and recorded participant responses. Auditory stimuli were broadband noise bursts gated on and off with a 15-ms rectangular window. Identical signals were presented from the two front or two rear loudspeakers. The perceived sound source location was manipulated by changing the level difference (panning stereophony) between either the front-left (L_F) and front-right (R_F) or between the back-left (B_L) and back-right (B_R) loudspeaker signals. Level differences of −5, 0, and 5 dB were used, with negative intensity ratios indicating that the left speaker signal was more intense than the right. This panning-based stereophonic technique generated a phantom sound source at ~ +/− 10 degrees and 0 degrees from both front and back. The average intensity for all auditory stimuli was maintained at 65 dBA, as verified using a sound level meter (Brüel & Kjær 2250-L) positioned at the location of a listener's head.

Custom-designed software written in Arduino controlled the visual stimuli. Visual stimuli were 15-ms duration blue LED light flashes (13.5 cd/m^2, measured at the LED surface) generated from one of three lighting positions. The onset of the light stimulus was synchronized with the onset of the sound stimulus at the leading loudspeaker as verified by oscilloscope measures.

2.3 Procedure

Two experimental conditions (auditory localization and visual localization) were conducted over five visits over five separate days. On days one and two, participants completed two auditory conditions followed by one visual condition. On days three through five, participants completed three auditory conditions followed by one visual condition. Each auditory and visual condition lasted approximately ten minutes and two minutes, respectively. Participants were given a three- to five-minute break period between conditions. Before beginning the experiment each day, participant eye tracking was calibrated using 13 known LED locations (from -12^o to +12^o at 2^o increments). This report focuses on the effects of vision on auditory localization; results for visual localization are not reported.

Before the experiment, to help participants understand the nature of the experiment and the response tasks expected, a training block for auditory localization was presented. Participants were presented with a sound from one of six locations (three in front and three in back) and were instructed to respond by 1) directing their gaze to the lateral direction of a sound and 2) pressing a button to indicate front or rear origin of the sound source. Participants were provided with alert feedback, following each erroneous front-back judgment, to aid in the proper performance of the front/back decision-making portion of the task. The training block continued until participants achieved >75% accuracy and felt ready to start an experiment.

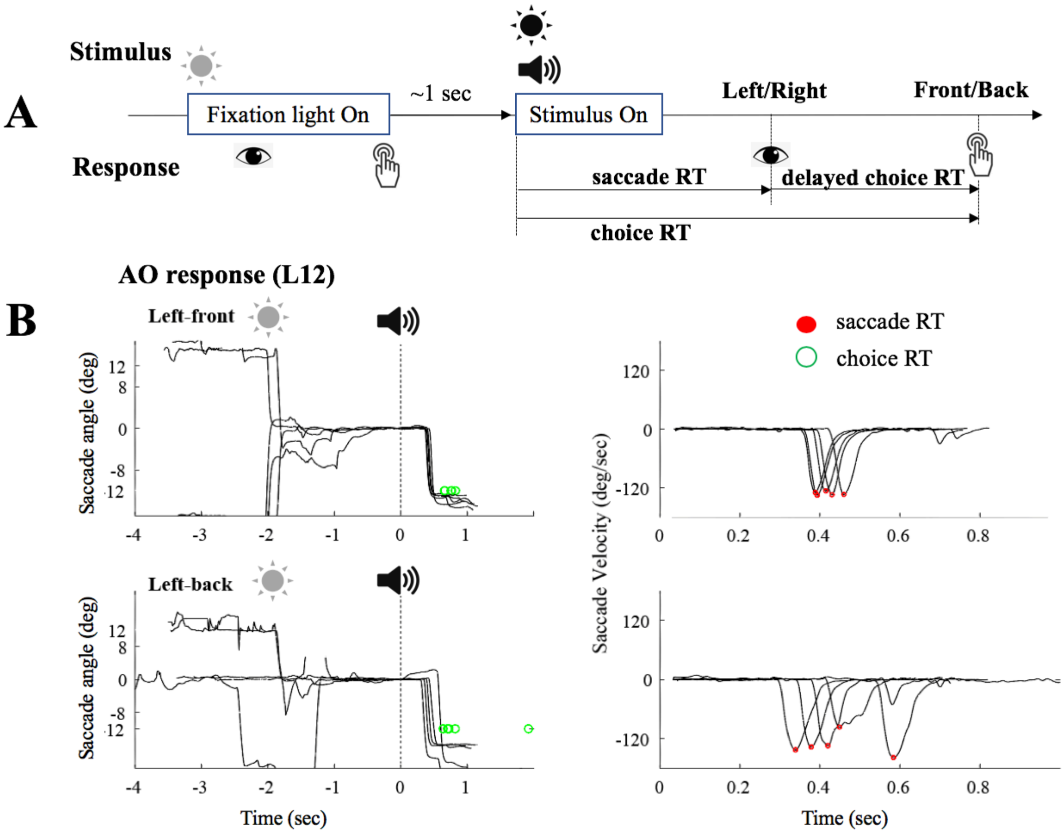

Figure 1 – Experimental procedure and reaction time analysis. (A) Stimulus-response sequences for each trial in the auditory localization task. (B) Example saccade and choice responses in response to auditory-only stimuli. Left, saccade angles to left-front and left-back sounds. Right, saccade velocity to these two stimuli. Light gray indicates the start time of center fixation light, the stimulus onset is marked as time zero.

2.4 Auditory Localization Task

Sound stimuli with or without lights were presented in randomized blocks, denoted as audio-visual (AV) and auditory-only (AO) blocks, respectively. The AO block contained six stimuli total (three in front and three in back). For the AV block, each of the six auditory stimuli was paired with one of the three LEDs, resulting in a total of 18 stimuli. The order of blocks and stimuli was randomized. Five repeats were administered for each stimulus, resulting in a total of 120 trials. For all trials, participants were instructed to indicate only the direction of the sound they heard, not the light they saw.

Figure 1A shows the stimulus and response sequences within a trial. To initiate a trial, a participant moved her/his gaze towards a red LED cue presented straight ahead at 0^{o}. The participant then pressed a button to indicate fixation and initiate stimulus presentation. While the participant maintained center fixation, the fixation light turned off. After ~1 second, a sound stimulus (AO block) or a pair of sound and light stimuli (AV block) were presented. The participant indicated the perceived location of the sound stimulus in two sequential steps: 1) rapidly directing his/her gaze toward the perceived direction of the sound source on the azimuth plane (saccade response) and 2) after the gaze shift, indicating the perceived front/back location of the sound by pressing a front/back button (choice response). Eye movements were recorded using an eye tracking system (90-Hz sampling rate, Arrington Research) and button presses were recorded using a modified gaming console, two buttons for front and back choices were arranged vertically on the game console. Participants were not provided any feedback or knowledge of their results during or after the experiments. They were unaware of the total number of loudspeakers and spatial location of each speaker.

2.5 Data Analysis

Figure 1B shows typical saccade and choice responses. Rapid saccade movements can be seen upon hearing a sound. The saccade endpoints (with a duration of ~15 ms) before a participant made a front/back choice response (green dot shows the choice RT) were averaged and used as the perceived lateral position of the sound source. The saccade velocity of eye movement after sound onset was then calculated. The time for achieving the peak velocity (red dot) was used as the saccade RT. Since the responses to lateral and front-back dimensions were made sequentially, the data analysis extracted the delayed choice RT (i.e., the difference between choice RT and saccade RT) to analyze the separate effects, if any, of vision on front-back decisions. The FBC errors were calculated by the percent of incorrect button-pressing choices a participant made when a sound was from the front but responses were to the back or vice-versa.

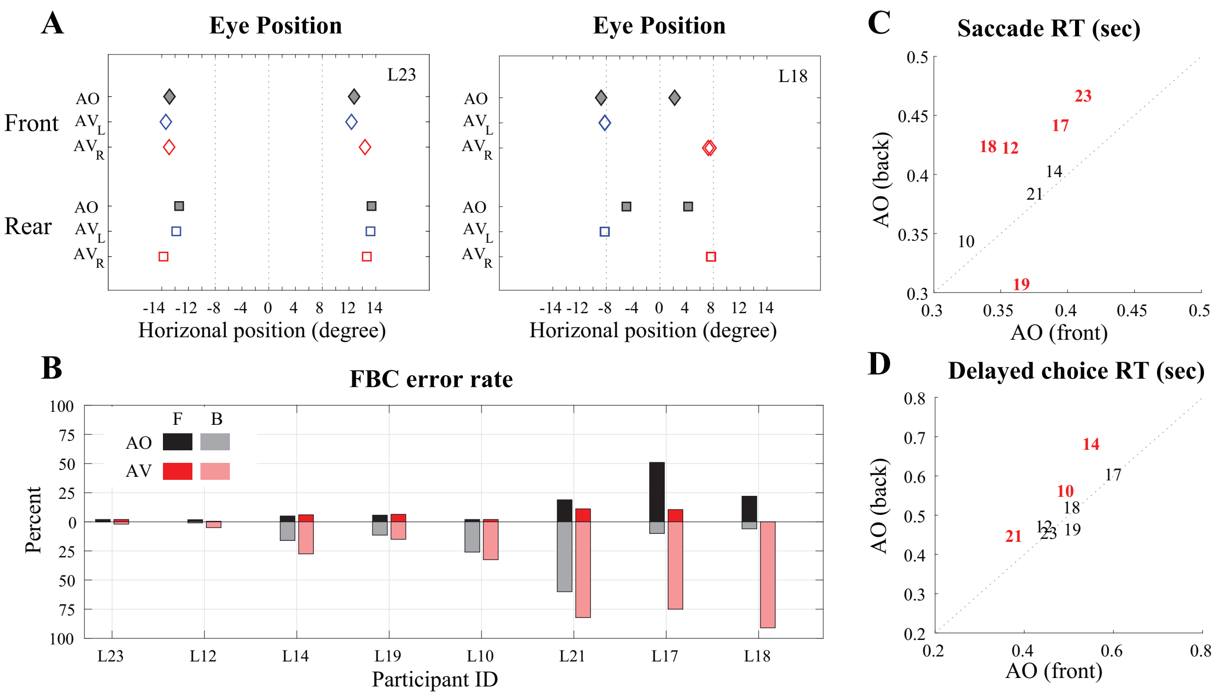

Figure 2 – Auditory localization responses. (A) Saccade-based lateral judgment of sound direction with and without light. Left and right responses are associated with the panned directions of left and right sound sources, respectively. (B) Averaged FBC error rate of all participants in AO and AV conditions. (C) Saccade RT and (D) Delayed choice RT for all participants in the AO condition. Red symbols (Participant ID) indicate significant differences between RT to front and rear sound (two-tailed *t*-test, $p<0.05$).

3. RESULTS

3.1 Response Accuracy

Results show that saccade eye movement can reliably indicate the perceived lateral direction of a sound source in both frontal and rear space. Even with the presence of light stimuli, the performances of five participants revealed little change in their saccade-based lateral judgment of the sound source direction. Figure 2A (left panel) shows one of these participants. On the other hand, two participants showed complete visual capture by light (Fig. 2A, right panel); their responses in the AV condition followed the light direction, dramatically shifted from their responses in the AO condition. This occurred despite repeated instructions to localize the sound and not the light. The remaining participant showed incomplete visual capture (not shown).

Individual variability is also observed in front-back judgments. Figure 2B shows the FBC rate for each of four stimulation types (sounds in front/back with or without light, 2×2 = 4) for all participants. The five participants showing limited visual influences on lateral judgments also reported limited errors for FBCs in both AO and AV conditions. Their data is shown in the left five columns. By contrast, the three participants showing strong visual capture for lateral judgments reported elevated incidences of FBC errors. In particular, light stimulation caused strong back-to-front response reversals for rear auditory targets, resulting in large FBC error rates (pink).

3.2 Reaction Time

All but one participant showed longer choice RT to sound stimuli presented from the rear than from the front. However, since front-back button pressing responses were made after saccades, this delay could have occurred during the saccade. Thus, we analyzed separately the saccade RT (Fig. 2C) and the delayed choice RT (Fig. 2D); see Fig. 1A for Methods. The results reveal that four participants showed longer saccade RT to rear sounds (red in Fig. 2C), but no difference in their delayed choice RT between front and back sounds (black in Fig. 2D).

On the other hand, three participants showed longer delayed choice RT to rear sounds (red in Fig. 2D), but no difference in their saccade RT to front and rear sounds (black in Fig. 2C). For the participant who reacted faster to rear sound (L19), it is the saccade RT, not delayed choice RT, that showed a significant difference between front and rear. These results suggest that the participants we tested made their front/back decision either before or after their saccade responses. For the former group, they likely needed to realize the actual position of a sound source before making a saccade to its lateral direction, whereas for the latter group, the left-right and front-back decisions were likely made sequentially, in part, due to experimental instruction. This difference might reflect a location-based vs. hemifield-based localization strategy among participants.

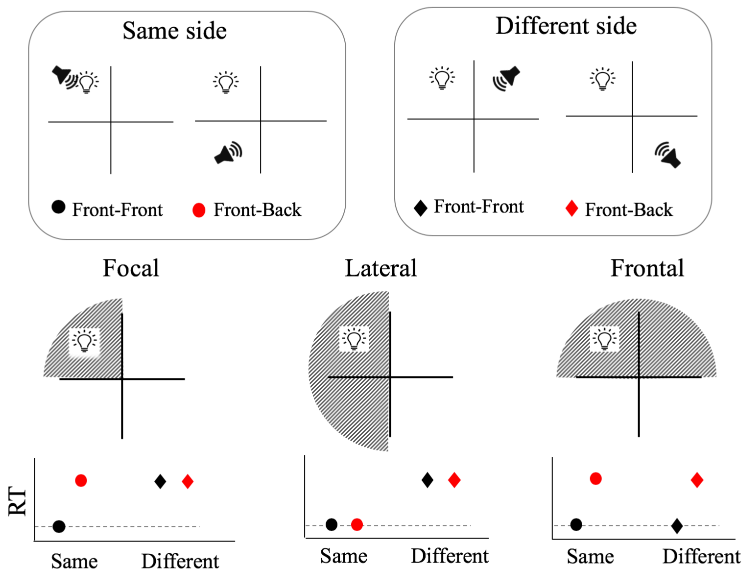

Figure 3 – Hypothesis for changes in RT with A-V congruent and incongruent stimulus conditions.

The existing literature shows that when sound and light stimuli are both in the frontal field, AV incongruent stimuli (on different sides) have longer RTs than AV congruent stimuli (on the same side) for auditory localization tasks (8, 9). However, little is known about the spatial extent of visual influence, or more likely, the extent of visual spatial attention, when a sound may come from either frontal or rear space. We hypothesize that three scenarios may be possible for panoramic auditory localization, as shown in Figure 3. (1) Visual-spatial attention is focal: any AV incongruity between left-right and/or front-back dimensions delays auditory response. (2) Visual-spatial attention is lateral: AV incongruity in the left-right dimension delays auditory response, independent of perceived front-back directions. (3) Visual spatial attention is frontal: AV incongruity in the front-back dimension delays auditory response, independent of perceived left-right direction.

The results show that although visual stimulation did not affect left-right auditory localization accuracy in some participants (N=4), their saccade RT to rear sound sources in AV response were significantly longer than those in AO responses (one-tailed *t*-test, $p<0.05$) and visually induced saccade delay was more apparent for left-right incongruent than congruent AV stimuli (Scenario 2). Figure 4A and 4B show the changes in RT between AV and AO conditions of two participants. Interestingly, delayed choice RT either showed no change or did not reveal this asymmetry. These results suggest that, when visual stimuli do not affect auditory localization responses (in terms of accuracy), visual spatial attention is lateral and acts fast for a brief period time before the saccade.

On the other hand, for participants (N=3) that experienced a great extent of visual capture of response accuracy, the change in their RT follows the pattern for the frontal visual spatial attention (Scenario 3) but in the opposite direction. That is, adding visual stimulation sped up their responses for rear stimuli. Figure 4C and 4D show the changes in RT between AV and AO conditions of two participants. The asymmetry between congruent and incongruent conditions also applied; incongruent front-back AV stimuli resulted in a significant change in RT. Interestingly, this asymmetrical change is reflected in either saccade RT (L18) or delayed choice RT (L21); one-tailed *t*-test, $p<0.05$. This suggests that when operating in the capturing mode, visual spatial attention is frontal and acts for a longer period time extending beyond lateral judgment through a saccade.

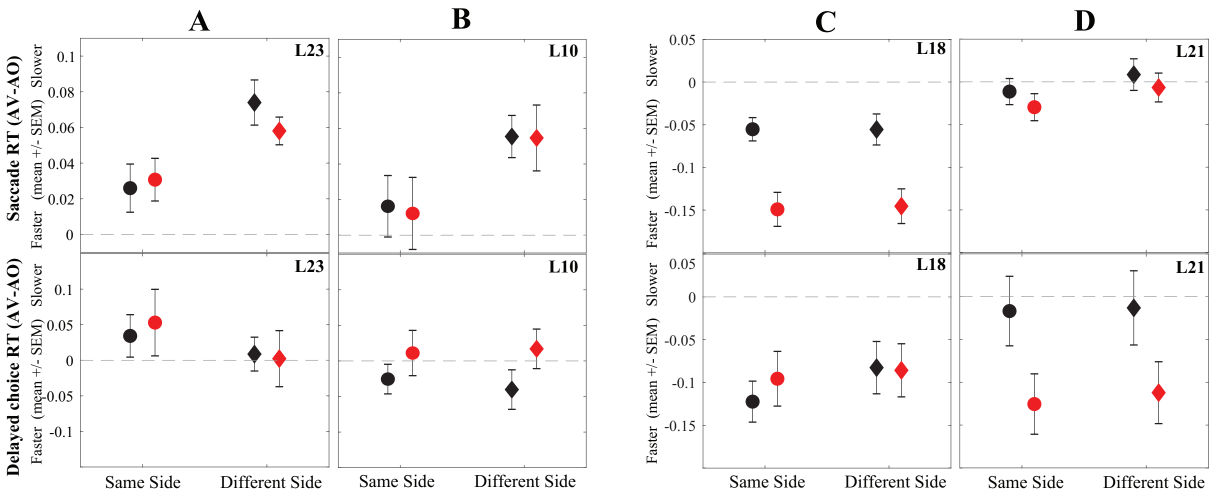

Figure 4 – Changes in saccade RT and delayed choice RT between AV and AO conditions for four participants. Black circle, sound is in front and on the same side of light; red circle, sound is in back and on the same side of light; Black diamond, sound is in front and from the opposite side of light; red diamond, sound is in back and from the opposite side of light. See Figure 3 for illustration of symbols.

4. CONCLUSIONS

The results of this study show that observed visual effects do not adhere to the spatial rule of multisensory interaction with regard to the physical proximity of cues. First, the influence of visual cues interacted closely with front-back confusions in auditory localization (Fig. 2B). Second, visual dominance in reaction time could extend to spatially incongruent auditory stimuli from the rear field, even in the absence of changes in response accuracy, e.g., L23 in Fig. 2A and Fig. 4A. When visual influences were weak and brief, the visual spatial attention channel appeared to operate in the lateral domain, delaying the auditory response from the other side. When visual influences were strong and sustained, the visual spatial attention channel appeared to operate in the front domain, attracting auditory responses from the back. These two different operating modes suggest a domain-specific difference in visual spatial attention between auditory and visual localization.

ACKNOWLEDGEMENTS

Christopher Montagne helped the hardware instrumentation and software development. Elise Wagener and Alina Roanne Lasrado contributed to the earlier versions of experimental design. We appreciate their help. This work is supported by NSF BCS-1539376 (YZ).

REFERENCES

1. Stein BE, Meredith MA. The Merging of the Senses. Cambridge, MA: MIT Press; 1993.

2. Jackson CV. Visual factors in auditory localization. Quarterly Journal of Experimental Psychology. 1953;5:52-65.

3. Pick HL, Warren DH, Hay JC. Sensory conflict in judgments of spatial direction. Perception & Psychophysics. 1969;6(4):203-5.

4. Thurlow WR, Jack CE. Certain determinants of the "ventriloquism effect". Perceptual and motor skills. 1973;36(3):1171-84.

5. Slutsky DA, Recanzone GH. Temporal and spatial dependency of the ventriloquism effect. Neuroreport. 2001;12(1):7-10.

6. Choe CS, Welch RB, Gilford RM, Juola JF. The “ventriloquist effect”: Visual dominance or response bias? Perception & Psychophysics. 1975;18(1):55-60.

7. Howard IP, Templeton WB. Human spatial orientation. New York: Wiley; 1966. 359-62 p.

8. Lukas S, Philipp AM, Koch I. Switching attention between modalities: Further evidence for visual dominance. Psychological Research PRPF. 2010;74(3):255-67.

9. Ragot R, Cave C, Fano M. Reciprocal effects of visual and auditory stimuli in a spatial compatibility situation. Bulletin of the Psychonomic Society. 1988;26(4):350-2.

10. Colavita FB. Human sensory dominance. Perception & Psychophysics. 1974;16(2):409-12.

11. Montagne C, Zhou Y. Audiovisual Interactions in Front and Rear Space. Front Psychol. 2018;9:713.

12. Carlile S, Leong P, Hyams S. The nature and distribution of errors in sound localization by human listeners. Hear Res. 1997;114(1-2):179-96.

13. Macpherson EA, Middlebrooks JC. Localization of brief sounds: effects of level and background noise. J Acoust Soc Am. 2000;108(4):1834-49.

14. Middlebrooks JC, Green DM. Sound localization by human listeners. Annual review of psychology. 1991;42(1):135-59.

15. Blauert J. Spatial hearing: The psychophysics of human sound localization. Revised edition. ed. Cambridge, MA.: MIT Press; 1997.