DOI: 10.61782/fa.2025.0674

FORUM ACUSTICUM EURONOISE 2025

PERCEPTION OF SIMULATED MOVING SOUND SOURCES IN REVERBERANT VIRTUAL AUDIO-VISUAL ENVIRONMENTS

Stefan Fichna[1*] **Steven Van de Par**[2] **Virginia L. Flanagin**[3] **Stephan D. Ewert**[1]

[1] Medizinische Physik, Universität Oldenburg, Germany
[2] Akustik, Universität Oldenburg, Germany
[3] Dt. Schwindel- und Gleichgewichtszentrum
LMU Klinikum München, Germany

ABSTRACT

Auditory distance perception plays a vital role in spatial awareness and navigation, particularly for estimating trajectories to avoid collisions with moving objects. With growing interest in virtual acoustics, simulating complex acoustic environments that mirror real-life scenarios has become increasingly important. While research has extensively examined static sound sources, less is understood about how distance is estimated when sound sources move. This study addressed this gap in a headphone-based experiment using a real-time room acoustics simulation enabling 6-degrees of freedom movement of source and receiver. A head-mounted-display (HMD) and a computer game engine were used to display the virtual visual environments and to conduct the experiment. We evaluated just noticeable distance thresholds and the perception of invisible frontal moving sound sources by estimating and visually indicating the start and end point of a linear motion. Stationary sound sources were additionally used to provide a comprehensive analysis of both static and dynamic conditions. Listeners detected the motion and perceived motion distance was greater in a room with longer reverberation time than in a less reverberant room. Our findings enhance our understanding of how humans process complex spatial information and have potential applications in virtual reality, and assistive technologies.

Keywords: *virtual acoustics, room acoustics simulation, distance estimation, moving sound source*

[*]***Corresponding author***: stefan.fichna@uol.de

1. INTRODUCTION

Auditory spatial perception, including the ability to localize sound sources, plays a fundamental role in how humans perceive and respond to their acoustic environment [1-2]. While static sound sources have been extensively studied, less attention has been given to moving sources. Accurately perceiving the distance of moving sound sources is crucial for real-world navigation, including an avoidance of collisions [3-6].

With advancements in virtual acoustic simulations, dynamic sound rendering has become an area of growing interest. Virtual audio-visual environments provide a controlled setting to explore auditory perception in complex scenarios [7]. Simulating realistic reverberation and movement patterns allows researchers to investigate how humans process spatial auditory cues. However, the interplay between reverberation, sound source motion, and perception in virtual environments remains underexplored. Most research on auditory distance perception has focused on static sources, investigating how intensity, spectral cues, and reverberation contribute to distance estimation [8].

Modern virtual reality (VR) technology and real-time room acoustics simulations enable precise manipulation of reverberation time, source movement, and listener dynamics [9]. Implementing six degrees of freedom (6-DOF) movement for both the source and receiver allows for detailed analysis of auditory motion perception in dynamic environments.

A central question is whether the amount of reverberation affects distance perception of a moving sound source. It appears that there is a lack of direct empirical studies specifically examining the effect of reverberation on perception of moved distance in auditory contexts. In assistive technologies, better understanding of auditory motion perception could improve the design of auditory navigation aids for visually impaired individuals [3, 10].
This study investigates auditory distance perception of moving sound sources in virtual environments with different levels of reverberation. We focus on (1) the estimation of the mean distance from the receiver to the sound source and (2) movement trajectory estimation. Using headphones, a room acoustics simulator, a head-mounted display and a computer game engine for visual rendering, we created realistic spatial listening conditions in a visually neutral grey room. This research advances our psychoacoustic knowledge of dynamic auditory cue processing in reverberant settings. The findings also could impact VR applications where accurate spatial representation is crucial for user experience and interaction fidelity.

2. METHODS

2.1 Listeners

Six normal-hearing listeners (two males and four female) aged between 17 and 28 years with a mean of 23.33 years and a standard deviation of 3.86 years participated in this study. Four of the six participants received hourly compensation. The other two listeners were employed by the University of Oldenburg. The participants did not report any hearing impairments and had either normal vision or vision that was corrected to normal using glasses or contact lenses.

2.2 Audio-Visual Environments

Two simulated, shoe-box shaped rooms were used in this study: A small room with a size of 5 x 8 x 2.5 m and a reverberation time RT60 of 0.4 s across all frequencies and a large room with a size of 10 x 16 x 4 m with a reverberation time RT60 of 1.4 seconds across all frequencies. In both rooms, the participant was positioned at the middle of the short wall (2.5 m in the small room and 5m in the large room) and 1.5 m away from the short wall at one end of the room (see Fig. 1). An additional anechoic acoustic environment was generated for pre-measurements excluding all reflections and just taking direct sound into account.

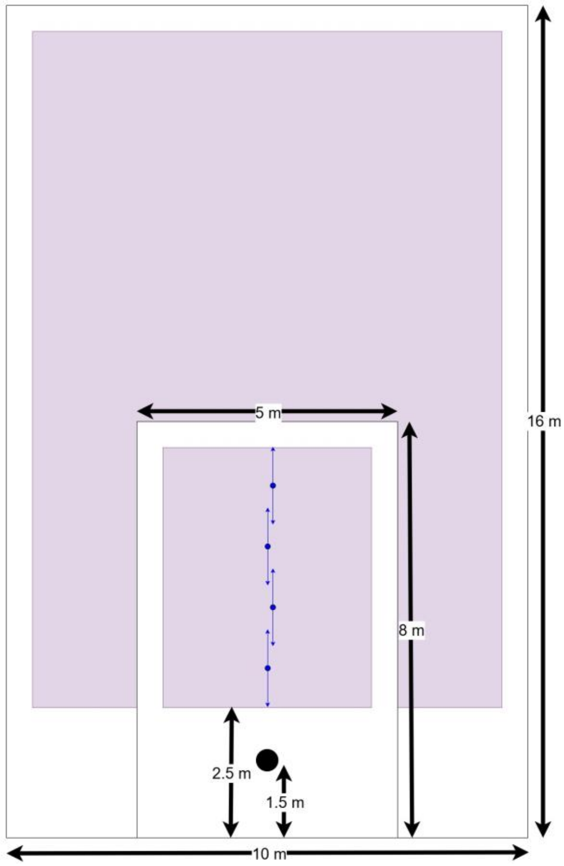

Figure 1. Dimensions and arrangement of receiver and sources in the small room and the large room. The black dot indicates the position of the participant. The blue dots and arrows show the possible motion center points and motion distances for the sound source with slight horizontal offsets for improved visibility. The purple colored areas represent the area in which participants could mark the perceived positions of the sound source.

The three auditory environments were created using the room acoustics simulator RAZR [11-12] (freely available at www.razrengine.com). RAZR calculates early reflections up to the third order using an image source model [13], while late reverberation is computed using a feedback delay network [14]. An assessment of various room acoustic parameters and subjective ratings of perceived room acoustical attributes demonstrated a good correspondence between simulated and real rooms [11; 15-16]. For the study presented here, the real time version liveRAZR was used, enabling dynamic 6-DOF source and receiver motion with 6-DOF.

For the visual component, rendered via a head-mounted-display (HMD), as well as for the control and execution of the experiment, Unreal Engine 4.27 was used. Sound source and receiver positions were sent in real-time via open sound control (OSC) messages to liveRAZR. A gaming controller, used with the right hand allowed the participant to interact with the environment and to provide responses.

The visual rooms contained no objects and were empty. The walls and the ceiling were flat and white. The part of the floor surface which is marked in purple in Fig. 1, consisted of a cracked stone floor. This surface was chosen to provide an additional visual distance and scale cue.

2.3 Stimuli and source motion

The stimuli for all measurements were a train of pink-filtered noise bursts (six bursts per second), each with a duration of 30 ms, gated using a Hann window to ensure smooth onset and offset transitions.

In the main experiment, an invisible sound source spawned at a starting point in the room, shown as blue dots in Fig. 1, at the height of the listener's head. All positions were directly in front of the participant. As the starting point is located in the middle of the motion trajectory, this point is referred to as the motion center point in the following. Overall four different motion center points were used with a distance to the participants of 1.75 m, 2.92 m, 4.09 m and 5.25 m. These distances were selected so that the full depth of the small room was utilized.

From the starting position there were three possible ways, how the sound source moved through space. A) The sound source moved in a sinusoidal fashion (maximum speed of 0.8 m/s) first towards the participant, back to the motion center point (away from the participant) and again back to the motion center point, were the motion terminated and the pulse train ended. B) The sound source moved the opposite way and first moved away from the participant and afterwards towards the participant. C) The sound source stayed stationary.

The distance between the turning points of the sound source motion was 1.5 m, referred to as motion distance in the following. The same movements with the same distance to the participants were simulated in both, the large and small virtual room. The overall motion and signal duration was six seconds. The signals were calibrated to ensure that the target signal level was 61 dB SPL at the furthest point of motion, located 5.25 meters from the source, within an anechoic environment.

For a pre-measurement of the just-noticeable difference (JND) between two stationary sound source distances, a similar but shorter one second long signal was used.

2.4 Apparatus and measurement procedure

Participants were seated in a soundproof booth with double walls, wearing Sennheiser HD 650 headphones connected to an RME Fireface UCX audio interface running at a sampling rate of 44.1 kHz. All listening tests were conducted using Matlab. For visualization, the Valve Index stereoscopic HMD by Valve Corporation (Bellevue, WA, USA) was utilized [17] together with the controller for the right hand to carry out the test.

For measuring the static distance JND, a two-alternative, forced-choice (2AFC) method was used with adaptively varying distance offset, in the small and the large room, as well as an anechoic room. Visually, the previously described rooms were used, with the difference that in both rooms for this measurement there was a blue colored line along which the invisible sound sources could spawn. For the anechoic measurement, a completely black virtual environment and also the blue-colored line for orientation were displayed. One sound source was positioned at a distance of 3.5 m minus a variable offset, while the other was positioned at 3.5 m plus the same offset. The order of presentation (i.e., whether the closer or farther sound source was presented first) was randomized across trials. Participants were instructed to indicate whether the second sound source appeared closer or farther than the first by moving a joystick on the controller up or down. The participants received feedback whether the answer was correct or not. The variable offset was adjusted adaptively until the participant could no longer reliably discriminate the distance between the two sound sources. Three reversals of the adaptive track were performed until a single threshold measurement was completed. JNDs were measured twice in a test and twice in a retest with the test being conducted at the start of the first appointment and the rest being conducted at the end of the second appointment.

The main measurement on the perceived motion trajectory was conducted after the first JND measurement.

After listening to a moving or stationary sound source as described in section 2.3, the participants positioned the visible loudspeakers, which were mounted on a pole at head height, by using the controller. In the virtual world, the controller had a laser pointer attached to it which allowed the participant to position two loudspeakers at the position, where the laser pointer touched the floor.

The participants were asked to position the visible loudspeaker to the perceived closer and farther turning point of the (invisible) sound source trajectory. The area in which the participant could position the loudspeakers is marked in purple in Fig. 1. The participants were able to change the position of the speakers as often as they liked before confirming their final answer with another button press. Afterwards, the next presentation started.

During the presentation, the participants sat on a chair. They were not allowed to move around the room but were permitted to move their head. The main measurements were conducted in two sessions. Each measurement began with written instructions, followed by a ten-minute familiarization phase in the small and large room with a visible sound source represented by the loudspeaker later used for distance estimation. In the second part of the familiarization the sound source was not visible. During the entire familiarization phase, the participants received feedback after estimating the motion turning points: Two green colored loudspeakers spawned at the positions between which the sound source actually moved. During the familiarization and main measurement, conditions were included in which the sound source was shifted to the left or to the right by one meter and where the sound source moved with a radial motion or a larger motion distance. These conditions were added to convey to the test subjects that the sound source can be located at any position in the room and can move in a wide variety of ways. This was also communicated to them in the written instructions. Of the six participants, three began in the first measurement block of the first session with a measurement in the small room, followed by a measurement block in the large room. The order was reversed for the other three participants.

3. RESULTS

The distance JND was analyzed by calculating the average across all participants. In the anechoic environment, the distance JND was 10.64 cm corresponding to a level JND of 0.53 dB. The distance JND was higher in the small room with a mean of 15.78 cm (level JND of 0.78 dB) and was

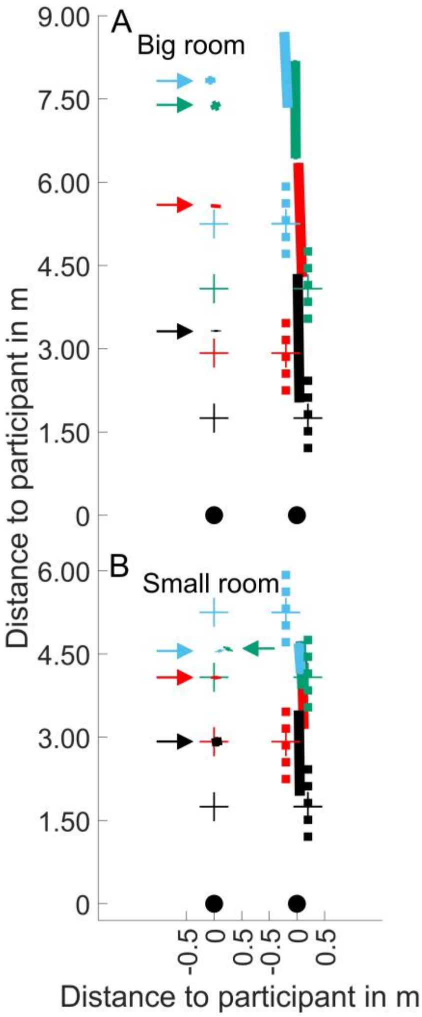

Figure 2. Top-down view of the motion distance estimation results. The x- and y-axis indicate the distance to the participant (black dot) in meters. The upper and lower parts show the results in the large room (A) and small room (B), respectively. On the left, results without source motion and on the right side results with 1.5 m source motion are shown. Crosses in black, red, green, and blue indicate the correct motion center points of the sound source. The dotted lines indicate the corresponding motion distances, with slight horizontal offset for improved visibility. Solid lines and dots show average results over all participants. The arrows point to the results without source motion.

the highest in the large room with 20.32 cm (level JND of 1.01 dB).

Figure 2 shows the result for the perceived motion trajectories as a top-down view onto the virtual environment. Section A presents the results obtained in the large and section B the results in the small room, respectively. The left side illustrates conditions in which the sound source was stationary, whereas the right side represents conditions where the source moved over the distance of 1.50 m. The x- and y-axes indicate the distance to the participant, whose position is marked by the black dot. Crosses denote the four motion center points at distances of 1.75 m (black), 2.92 m (red), 4.09 m (green), and 5.25 m (blue) from the participant. In conditions involving source motion, dashed lines indicate the actual motion distance. To improve readability, the lines have been slightly offset horizontally, although during the experiment, all motions were on a line directly in front of the participant. Solid lines and dots represent the mean response across all participants, indicating where participants placed their two markers. The colors of these markers correspond to the colors of the motion center crosses. In conditions without source motion (left side), the response markers were closely clustered, making individual data points difficult to distinguish. Therefore, arrows have been added to indicate these points. In general, the motion center point tended to be overestimated in trials involving shorter distances (1.75 m, 2.92 m, and 4.09 m). This overestimation is particularly evident on the left side of Fig. 2, where responses for the 1.75 m condition (black dot) align with or extend beyond the actual motion center point at 2.92 m (red cross), and responses for the 2.92 m condition (red dot) are similarly at or beyond the correct motion center point at 4.09 m (green cross). This overestimation effect appears to diminish at greater distances. In the small room (Fig. 2A), the motion center point at 5.25 m (blue cross) was then underestimated. The solid lines in the right part of the figure show that the estimated motion distances exhibit considerable overlap, with a similar estimated motion center point as for the conditions without source motion. It is evident that the perceived motion distance is significantly greater in the large room than in the small room.

Figure 3 illustrates the estimated motion distance for each of the four motion center points in the large room (section A) and the small room (section B). The blue violin plots represent trials in which the sound source was stationary (motion distance = 0 m), while red violin plots correspond to trials in which the motion distance was 1.5 m. The correct motion distance is indicated by a solid horizontal line, while dotted horizontal lines mark 0 m (no motion

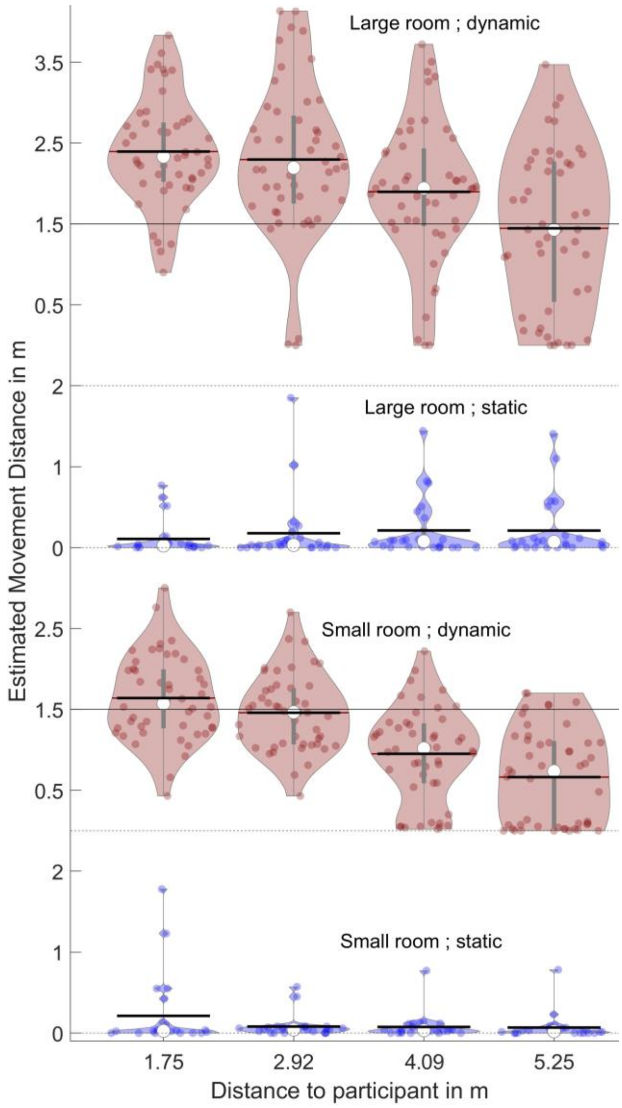

Figure 3. Estimated motion distances for all four motion center points indicated on the x axis. The upper part (A) shows results in the large room and the bottom part (B) shows the results in the small room. Red and blue violin plots indicate a motion distance of 1.5 m and 0 m, respectively. The solid horizontal grey lines indicate the correct motion distance of 1.5 m for the red plots. Each violin plot includes darker dots for the individual responses of the participants. The white dot shows the median and the black line the average across all participants.

distance). The individual responses are shown as dots in the violin plots. The white circle denotes the median, and the black line indicates the mean value.
The blue violin plots show that participants generally recognized when the source remained stationary. However, occasional outliers suggest that some participants perceived motion distances of close to 2 m, an effect observed in both the large and small room. The red violin plots in Fig. 3A for trials conducted in the small room indicate that at shorter distances, the estimated motion distance closely approximated the actual motion distance (solid horizontal line). However, for larger distances, the motion distance tended to be underestimated. This underestimation is partly due to the fact that, at greater distances, participants frequently failed to perceive any motion at all. Overall, the estimated motion distance decreases as the motion center point moves farther from the participant.

4. DISCUSSION

4.1 JNDs for distance estimation

The results indicate that the JND in auditory distance perception increases with reverberation time. Participants exhibited the lowest JND in the anechoic room, followed by the small reverberant room, with the highest JND observed in the large reverberant room. This finding aligns with previous research demonstrating that reverberation introduces ambiguity in distance cues, particularly when level and direct-to-reverberant energy ratio cues are diminished [8]. The JNDs found in this study align with findings from Mills (1960) and Klockgether & van de Par (2016) [18-19]. The congruence between our findings and those reported in these studies supports the validity of our measurement system.

4.2 Auditory distance perception

Systematic biases in distance perception were found in this study across different room conditions. In the small room, participants generally overestimated the position of the sound source for the closest three motion center points (1.75 m, 2.92 m, and 4.09 m). However, the farthest point at 5.25 m was consistently underestimated (see Fig. 2B). These results are broadly consistent with compression in auditory distance perception [2; 8]. A possible explanation for these observations lies in the near-field and far-field distinction. In the near field (closer sound sources), auditory cues such as changes in the direct-to-reverberant energy ratio are stronger and more precise, facilitating more accurate distance estimation. As the source moves into the far field (greater distances), these cues become weaker and more ambiguous, increasing the likelihood of underestimation [8]. This effect was particularly apparent in the small room, where participants underestimated the motion center point at 5.25 m.
In the large room, an overestimation of the distances was observed consistently across all motion center points. This may be due to a perceptual scaling effects arising from the visual impression of the room. Participants may have implicitly adapted their distance estimations to the larger room and attempted to “fill” the available space, resulting in consistent overestimations. The longer reverberation time in the large room likely contributed to the blurring of spatial boundaries and an expansion of the perceived auditory image[20]. Such contextual scaling effects may be especially prominent in virtual environments where visual anchoring is limited.

4.3 Perceived motion distance

The current results suggest that the auditory perceived motion distance of an invisible sound source is systematically influenced by room acoustics and size of the visual environment. Participants generally overestimated moved distances in the large reverberant room compared to the small room (see Fig. 2), consistent with the effects observed for auditory distance perception. This supports the hypothesis that longer reverberation time leads to an expansion of perceived space, as reflections extend the auditory image and blur motion boundaries [20-21]. Previous studies have suggested that reverberation can enhance spatial envelopment and source width [1; 7], which may contribute to uncertainty and an exaggerated perception of motion distance observed in this study.
Additionally, we observed that in the small room, perceived motion distances were more accurate at closer distances but underestimated at farther distances, as shown in Fig. 3B. This result aligns with findings from Kolarik et al. [10], who reported that auditory motion perception is more reliable for near-field sources due to stronger binaural and spectral cues. At greater distances, the weakening of these cues and increased reliance on reverberation may contribute to the observed underestimation effect. While no underestimation was observed for the larger room, the same trend of smaller estimated motion distances for farther motion center points was observed. To estimate a potential effect of the virtual pointing method on the results, a control condition with indicating distance of visual objects at the sound source location should be included.

4.4 Stationary vs. moving sound sources

Participants reliably distinguished stationary sources from moving sources in both room conditions. However, occasional outliers in responses to stationary sources suggest that some participants perceived illusory motion. This phenomenon has been reported in previous studies, where reverberant reflections can create an impression of movement even when the source remains static [23]. The presence of these misperceptions underlines the complexity of auditory motion processing and the challenges associated with accurately interpreting dynamic acoustic cues in reverberant environments.

5. CONCLUSION

This study demonstrates that reverberation significantly influences auditory distance perception and motion trajectory estimation in virtual acoustic environments. The results indicate that longer reverberation times lead to larger perceived motion distances. Furthermore, systematic biases were observed in distance perception: While participants generally underestimated sound source distances in the small room at closer positions, they overestimated the farthest point. In the large room, overestimations occurred consistently across all positions, which we attribute to a perceptual scaling effect resulting from the room's larger spatial extent in the visual domain.
Future research should more explicitly explore the effect of binaural cues and motion parallax for laterally displaced sound sources and for a larger variety of motion distances.

6. ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft, DFG SPP Audictive – Project-ID 444827755.

7. REFERENCES

[1] Blauert, J. (1997). Spatial hearing: The psychophysics of human sound localization (Revised edition). MIT Press

[2] Zahorik, P., Brungart, D. S., & Bronkhorst, A. W. (2005). Auditory distance perception in humans: A summary of past and present research. Acta Acustica united with Acustica, 91(3), 409–420.

[3] Steffens, H., Schutte, M., & Ewert, S. D. (2022). Acoustically driven orientation and navigation in enclosed spaces. *The Journal of the Acoustical Society of America*, *152*(3), 1767-1782.

[4] Schutte, M. (2021). Aspects of room acoustics, vision and motion in the human auditory perception of space (Doctoral dissertation, LMU).

[5] Rosenblum, L. D. (2010). Perceiving movement and location through auditory perception. Ecological Psychology, 22(4), 278–289. https://doi.org/10.1080/10407413.2010.510019

[6] Oberfeld, D., Wessels, M., & Huisman, T. (2023). Auditory and audiovisual time-to-collision estimation and road-crossing decisions.

[7] Begault, D. R. (2000). 3D Sound for Virtual Reality and Multimedia (Academic Press, New York).

[8] Zahorik, P., & Wightman, F. L. (2001). Loudness constancy with varying sound source distance. Nature Neuroscience, 4(1), 78–83. https://doi.org/10.1038/82923

[9] Lindau, A., & Weinzierl, S. (2012). Assessing the plausibility of virtual acoustic environments. Acta Acustica united with Acustica, 98(5), 804–810. https://doi.org/10.3813/AAA.918560

[10] Kolarik, A. J., Cirstea, S., Pardhan, S., & Moore, B. C. (2013). A summary of research investigating echolocation abilities of blind and sighted humans. Hearing Research, 310, 60–68. https://doi.org/10.1016/j.heares.2014.01.010

[11] Wendt, T., Van De Par, S., & Ewert, S. D. (2014). A computationally-efficient and perceptually-plausible algorithm for binaural room impulse response simulation. Journal of the audio engineering society, 62(11), 748-766, doi: 10.17743/jaes.2014.0042.

[12] Kirsch, C., Wendt, T., van de Par, S., Hu, H., and Ewert, S. D. (2023). "Computationally-Efficient Simulation of Late Reverberation for Inhomogeneous Boundary Conditions and Coupled Rooms," Journal of the Audio Engineering Society, 71, 186–201. doi:10.17743/jaes.2022.0053

[13] Allen, J. B., & Berkley, D. A. (1979). Image method for efficiently simulating small-room acoustics. The Journal of the Acoustical Society of America, 65(4), 943-950, doi: 10.1121/1.382599.

[14] Jot, J.-M., & Chaigne, A. (1991). Digital delay networks for designing artificial reverberators. Paper presented at the Audio Engineering Society Convention 90.

[15] Brinkmann, F., Aspöck, L., Ackermann, D., Lepa, S., Vorländer, M., & Weinzierl, S. (2019). A round robin on room acoustical simulation and auralization. The Journal of the Acoustical Society of America, 145(4), 2746-2760, doi: 10.1121/1.5096178.

[16] Blau, M., Budnik, A., Fallahi, M., Steffens, H., Ewert, S. D., and van de Par, S. (2021b). "Toward realistic binaural auralizations – perceptual comparison between

measurement and simulation-based auralizations and the real room for a classroom scenario," Acta Acust., 5, 1–13. doi:10.1051/aacus/2020034

[17] Angelov, V., Petkov, E., Shipkovenski, G., & Kalushkov, T. (2020). Modern virtual reality headsets. Paper presented at the 2020 International congress on human-computer interaction, optimization and robotic applications (HORA), doi: 10.1109/HORA49412.2020.9152604.

[18] Mills, A. W. (1960). On the minimum audible angle. The Journal of the Acoustical Society of America, 32(1), 132–134. https://doi.org/10.1121/1.1907864

[19] Klockgether, S., & Van de Par, S. (2016). Auditory distance perception in reverberant and anechoic environments. The Journal of the Acoustical Society of America, 140(3), 1833–1842. https://doi.org/10.1121/1.4962562

[20] Larsson, P., Västfjäll, D., & Kleiner, M. (2010). Perception of sound source motion in virtual environments: Effects of motion velocity and reverberation. The Journal of the Acoustical Society of America, 128(6), 3032–3043. https://doi.org/10.1121/1.3493465

[21] Shinn-Cunningham, B. G. (2000). Distance cues for virtual auditory space. In Proceedings of the IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (pp. 16–19). IEEE.

[22] Brungart, D. S., & Simpson, B. D. (2002). Effects of temporal fine structure on the localization of moving sound sources. The Journal of the Acoustical Society of America, 112(1), 263–273. https://doi.org/10.1121/1.1482076

[23] Neuhoff, J. G. (2001). An adaptive bias in the perception of looming auditory motion. Ecological Psychology, 13(2), 87–110. https://doi.org/10.1207/S15326969ECO1302_2