forum acusticum 2026
8-12 September · Graz, Austria

INFLUENCE OF REVERBERATION ON THE HORIZONTAL LOCALIZATION OF SOUND SOURCES AURALIZED IN AN AUDIOVISUAL IMMERSIVE ROOM

J. Vaquerizo-Serrano[1], M. Arnela[1]
[1] *Human-Environment Research (HER) group, La Salle, Universitat Ramon Llull, Catalonia, Spain*

Abstract

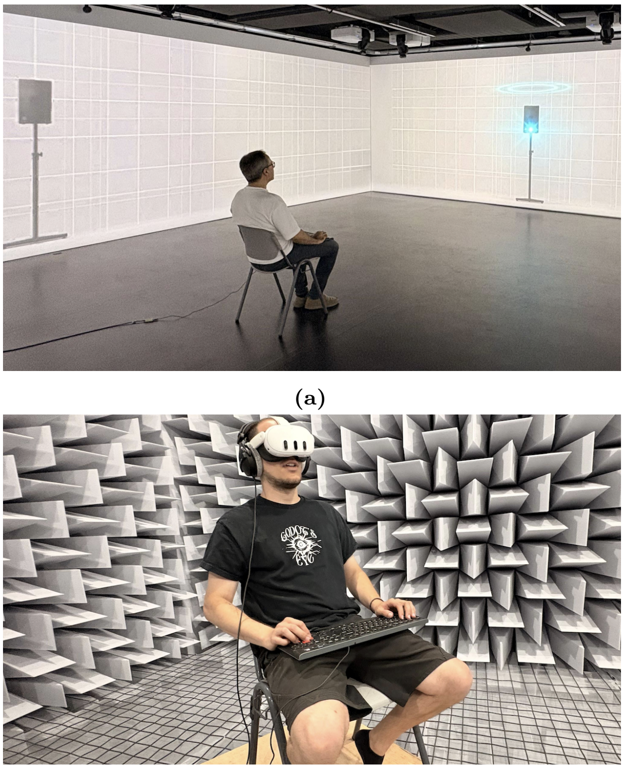
Audiovisual immersive rooms represent emerging virtualization spaces that address the individual isolation inherent to head-tracked binaural Virtual Reality (VR). However, auralizing highly reverberant environments remains a challenge, as reverberant energy degrades acoustic cues essential for human spatial localization. This study evaluates sound source localization in reverberant fields within a 16.2.10 audiovisual immersive room, comparing results to traditional head-tracked binaural VR. A 4-Alternative Forced Choice (4-AFC) horizontal localization task was conducted with 26 participants. Stimuli were auralized using pink noise bursts and Third-Order Ambisonics impulse responses generated in ODEON for the same reverberation room, with absorption progressively increased to produce reverberation conditions ranging from highly reverberant to nearly anechoic. As expected, the results demonstrate that localization accuracy decreased with increasing reverberation. However, across all conditions, the immersive room achieved higher accuracy and lower task response latency, exhibiting fewer localization confusions and lateral mislocalizations than binaural VR.

Keywords: *spatial audio, psychoacoustics, audiovisual immersive room, virtual reality, reverberation*

1. Introduction

The replication of realistic auditory environments has transitioned from static channel-based reproduction to interactive, spatialized systems [1], [2]. While binaural Virtual Reality (VR) offers exceptional portability, perceptual realism is often compromised. Non-individualized Head-Related Transfer Functions (HRTFs) introduce spectral distortions [3], leading to front-back confusion, lateral mislocalizations, and sound field internalization [4], [5].
Conversely, audiovisual immersive rooms recreate the sound field using loudspeakers distributed along the walls and ceiling, bypassing the above limitations by delegating spectral filtering to the listener's natural anatomy [6], [7]. While recent studies have attempted to stabilize VR externalization through augmented visual cues [8], immersive rooms inherently preserve spatial immersion and audiovisual congruency [9]. Evaluating how spatial localization degrades under reverberation is critical for validating these virtualization spaces [10]. Reverberant energy degrades the direct-to-reverberant ratio (DRR) and interaural cues [11]. In head-tracked binaural VR, injecting a dense reverberant tail through static, generic HRTF filters forces an artificial spectral fusion between direct sound and the diffuse field [12], [13].
In this work, we evaluate the horizontal spatial localization of sound sources in highly reverberant environments within a 16.2.10 audiovisual immersive room, comparing its performance to that of traditional head-tracked binaural VR (see Figure 1). The primary problem addressed is the extent to which the acoustic cues required for localization are preserved or degraded when varying levels of simulated reverberation are introduced into these two distinct reproduction paradigms. To avoid potential spatial artifacts from microphone-array capture [14], Third-Order Spatial Room Impulse Responses (SRIRs) were generated via geometrical acoustics software (ODEON). Twenty-six participants performed a 4-AFC localization task under four reverberation conditions using both reproduction modalities. Specifically, this study investigates whether the loudspeaker array of the immersive room offers a quantifiable advantage over binaural VR in terms of localization performance and response latency across all evaluated reverberant conditions.
The remainder of this paper is structured as follows: Section 2 describes the methodology, including stim-

Corresponding author: *marc.arnela@salle.url.edu.*



(a)

(b)

Figure 1: Participants performing the 4-AFC localization task in (a) the audiovisual immersive room and (b) in the binaural VR conditions.

ulus generation, the reproduction systems, and the perceptual evaluation protocol. Section 3 presents the experimental results concerning localization accuracy, spatial error patterns, and task response latency. Section 4 discusses the implications of these findings, and finally, Section 5 outlines the main conclusions drawn from the study.

2. Methodology

This section outlines the workflow executed to evaluate the spatial performance of the two auralization modalities. It encompasses the generation of the acoustic stimuli, their subsequent reproduction via loudspeakers and headphones, and the design of the perceptual listening test.

2.1. Stimuli Generation

The virtual acoustic fields were based on the reverberation chamber (see Figure 2) located at the Interactive Arts & Science Laboratory (IASlab), La Salle - Universitat Ramon Llull. Four sound sources were positioned at the cardinal directions (0°, 90°, 180°, and 270° azimuth), each at a distance of 1.50 m from the center and a height of 1.20 m. An acoustic model of this chamber was built in ODEON, incorporating the directivity data of the loudspeakers (d&b 5S) deployed in the physical space. The ODEON model was calibrated against in-situ measurements, ensuring

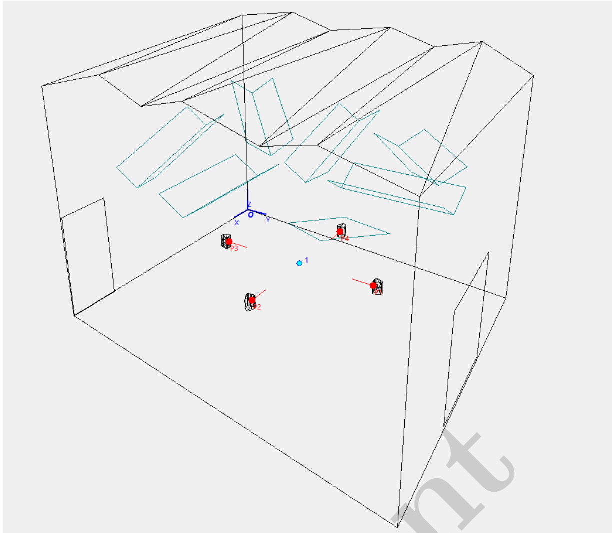


Figure 2: Acoustic model of the reverberation chamber built in ODEON for SRIR generation, showing the four simulated sound sources located at the cardinal positions.

that simulated acoustic parameters (T_{30}, C_{80} and T_s) remained within ±1 Just Noticeable Difference (JND). The ODEON model was adjusted to produce four distinct acoustic environments (see Table 1). The longest reverberation condition (RT3) was directly derived from the acoustic model, whereas RT1 and RT2 represent intermediate reverberation levels obtained by progressively increasing the absorption coefficients at the virtual room boundaries. RT0 corresponds to the quasi-anechoic reference condition with full absorption in all walls.

Third-Order SRIRs were generated from the ODEON model. Third-Order Ambisonics was selected as a practical compromise between spatial resolution and computational cost. To provide a controlled excitation signal, 240 ms broadband pink noise bursts were convolved with the generated SRIRs for each condition. This specific duration provides optimal excitation for horizontal localization tasks without inducing auditory fatigue during prolonged experimental sessions [5].

2.2. Immersive Room and Binaural VR Reproduction

Two different reproduction modalities were evaluated: an audiovisual immersive room based on a loudspeaker array (see Figure 1a) and a head-tracked binaural VR system (see Figure 1b). In both cases, the same HOA

Table 1: Reverberation conditions evaluated in the experiment and their corresponding reverberation times (T_{30}).

Reverberation Condition	T_{30} (s)
RT0	0.01
RT1	2.43
RT2	3.28
RT3	6.61

representations generated from the simulated SRIRs were used, allowing a direct comparison between two spatial-audio rendering approaches. The base stimuli were processed through a 1×16 spatial convolution with the SRIRs using the SPARTA suite [15]. This procedure preserved the HOA representation up to the final rendering stage, maintaining the spatial coupling and directional structure of the reflection pattern prior to rendering. Both spatial audio rendering pipelines were hosted within the REAPER Digital Audio Workstation (DAW).

For the immersive-room condition, the HOA signals were rendered through the 16.2.10 audiovisual immersive room using an Ambisonics decoding strategy implemented in SPAT Revolution and adapted to the Yamaha VXS8 loudspeaker layout. For the binaural VR condition, real-time head-tracking data from the Meta Quest 3 headset were captured and transmitted to the IEM SceneRotator plugin [16] to dynamically rotate the HOA sound field within REAPER. The rotated sound field was subsequently processed through the SPARTA Binaural Decoder plugin and reproduced using Beyerdynamic DT 770 PRO headphones.

Prior to the experiments, playback levels for both systems were matched to 65 dBA using a HEAD acoustics Head and Torso Simulator (HATS). Additionally, the system latency was measured, yielding 35 ms for the binaural VR setup and 658 ms for the immersive room audio pipeline.

2.3. Perceptual Evaluation

A total of 26 volunteers participated in the study (16 males, 10 females; age: $M = 37.0$ years, $SD = 11.3$, range = 23–57 years). All subjects reported normal hearing and had no history of auditory pathologies. A repeated-measures experimental design was adopted to evaluate the combined effects of reproduction modality and reverberation. The experiment consisted of 2 reproduction modalities $\times$ 4 reverberation conditions $\times$ 4 source positions $\times$ 3 repetitions, yielding a total of 96 trials per participant.

The execution of the perceptual experiment was entirely automated using a custom MATLAB application. This centralized system managed trial randomization and communicated with REAPER via Open Sound Control (OSC) messages to dynamically trigger stimulus playback across both reproduction modalities. Concurrently, MATLAB functioned as the core data acquisition system, logging participants' spatial responses and task response latencies for the subsequent statistical analysis.

Prior to the formal evaluation, participants completed a short familiarization phase using stimuli different from those employed in the main experiment. This training session allowed them to become acquainted with the 4-AFC localization task and response interface without prior exposure to the acoustic conditions evaluated during the study.

A counterbalanced experimental design was implemented to minimize order effects. Odd-numbered participants completed the immersive-room condition first followed by the binaural VR condition, whereas even-numbered participants performed the modalities in the reverse order. During the experimental phase, the auralized pink-noise bursts were presented from the four cardinal positions in a pseudo-randomized sequence. Participants remained seated throughout the experiment and were explicitly instructed to keep their head stationary and facing forward during stimulus presentation to prevent the use of dynamic localization cues.

For the immersive-room condition, the listening position coincided with the acoustic sweet spot located at the center of the loudspeaker array, and participants were aligned with the frontal reference (0° azimuth) before each block. In the VR condition, participants viewed a 360° stereoscopic photograph captured from the listening position in the immersive room, providing a congruent visual representation of the same environment, with a blue visual marker indicating the North direction embedded in the scene to anchor the frontal spatial reference.

Figures 1a and 1b show a participant performing the localization task in the immersive-room and binaural VR conditions, respectively. After each stimulus presentation, participants reported the perceived source location using a four-alternative forced-choice (4-AFC) response interface controlled via a numeric keypad, with one key assigned to each cardinal direction (North, South, East, and West).

3. Results

Figure 3 illustrates the evolution of localization accuracy across reverberation conditions. The immersive room accuracy declined from 94.0% in the quasi-anechoic condition (RT0) to 85.2% in RT3. In contrast, binaural VR exhibited a non-monotonic pattern, improving from 82.2% in RT0 to 85.2% in RT1 before decreasing to 78.5% in RT3. Although the immersive room achieved the highest overall accuracy, the performance gap between both modalities was largest in RT0 and became substantially smaller under reverberant conditions. Post-hoc comparisons revealed significant differences between RT0 and RT2 ($p = 0.010$) as well as between RT0 and RT3 ($p = 0.007$) within the immersive-room modality.

Figure 4 presents the confusion matrices aggregated across all participants for each condition. Responses in the immersive room remained strongly concentrated along the diagonal across all reverberation times, confirming its consistently high spatial accuracy. In contrast, binaural VR exhibited a higher overall error rate and a different error structure. Front-back confusions constituted the dominant source of error in the binaural condition, varying between 7.4% and 12.1% across reverberation levels, whereas lateral errors remained comparatively stable (7.4%–9.3%). These findings suggest that non-individualized binaural rendering primarily affected sagittal-plane discrimination, while

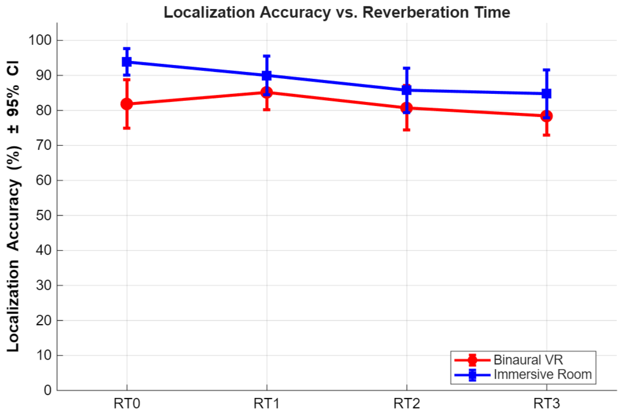


Figure 3: Mean localization accuracy as a function of reverberation condition for the immersive room and binaural VR.

the immersive room maintained a more robust spatial representation across all conditions.

Figure 5 shows the mean latency-compensated response times as a function of reverberation condition for both reproduction modalities. Participants responded significantly faster in the immersive room ($p < 0.001$). Statistical analysis revealed significant main effects of modality ($F(1, 25) = 91.59$, $p < 0.001$, $\eta_p^2 = 0.786$) and reverberation condition ($F(3, 75) = 28.85$, $p < 0.001$, $\eta_p^2 = 0.536$), but no significant Modality × RT interaction ($F(3, 75) = 0.89$, $p = 0.452$, $\eta_p^2 = 0.034$), indicating that the latency advantage of the immersive room remained relatively constant across reverberation conditions. These results suggest that increasing reverberation imposed an additional cognitive burden on the localization task regardless of the reproduction modality employed.

4. Discussion

The results demonstrate that the audiovisual immersive room yielded higher localization accuracy and shorter response latencies than binaural VR. However, the performance dynamics across acoustic environments revealed distinct behaviors for each reproduction modality. Statistical analysis indicated significant main effects of reproduction modality and reverberation condition, alongside a non-significant Modality × RT interaction for localization accuracy. From RT0 to RT3, accuracy decreased by 8.8 percentage points in the immersive room and 3.7 percentage points in binaural VR. Because the order of presentation was counterbalanced across participants, the observed modality differences cannot be readily attributed to learning or fatigue effects.

Although the immersive room provided a higher absolute baseline, its advantage over binaural VR was strictly condition-dependent. Specifically, the loudspeaker array outperformed the virtualized system by a significant margin of 11.8 percentage points in the anechoic baseline (RT0). Interestingly, as reverberation was introduced (RT1–RT3), this gap narrowed to approximately 5–6 percentage points and became statistically non-significant. The immersive room exhibited a progressive reduction in localization accuracy as reverberation increased.

The binaural VR showed a descriptive improvement from RT0 (82.2%) to RT1 (85.2%) before declining under the more reverberant conditions. Although this effect was not strong enough to produce a significant interaction, a possible explanation lies in the different perceptual roles of early and late reflections. As demonstrated by Gutiérrez-Parera et al. [17], early reflections can improve source localization and apparent source width, whereas late reverberation primarily contributes to envelopment. Consequently, the moderate early reflected energy introduced in RT1 may have provided supplementary spatial cues that temporarily enhanced directional discrimination and partially compensated for the limitations associated with non-individualized HRTFs [8], [12]. However, as reverberation density increased in RT2 and RT3, the direct-to-reverberant ratio (DRR) progressively decreased, reducing the salience of directional cues and increasing spatial blurring. Under these conditions, the detrimental influence of the late reverberant tail outweighed the potential benefits of the early reflections, resulting in degraded localization performance.

The confusion matrix analysis provides additional insight into the underlying mechanisms. In binaural VR, front-back confusions constituted the dominant source of error (peaking at 12.1% in RT3), while lateral mislocalizations remained comparatively stable, ranging between 7.4% and 9.3% across conditions. This aligns with classical literature stating that non-individualized spectral filtering primarily disrupts median-plane localization.

The immersive room successfully anchored the frontal focal point, keeping sagittal errors consistently lower, in contrast to what occurred with the binaural VR modality. Response latency results further support this discrepancy in perceptual effort. Participants consistently responded faster in the immersive room, maintaining a latency advantage of over 500 ms across all conditions. The highly significant effect of reverberation on latency, combined with a non-significant interaction, demonstrates that both modalities exhibited a parallel increase in response times as reverberation increased. The combination of lower accuracy and longer response times suggests that listeners required greater cognitive resources to resolve spatial uncertainty under reverberant virtualized conditions, although the acoustic blurring slowed down decision-making comparably in both systems.

Finally, it should be noted that the immersive room was evaluated under real-world operational conditions, where the real room response naturally coexisted with the virtualized sound field. Rather than representing a limitation, this reflects the operational context of real-world immersive audiovisual environments. Consequently, the present results provide insight into the practical performance differences between loudspeaker-

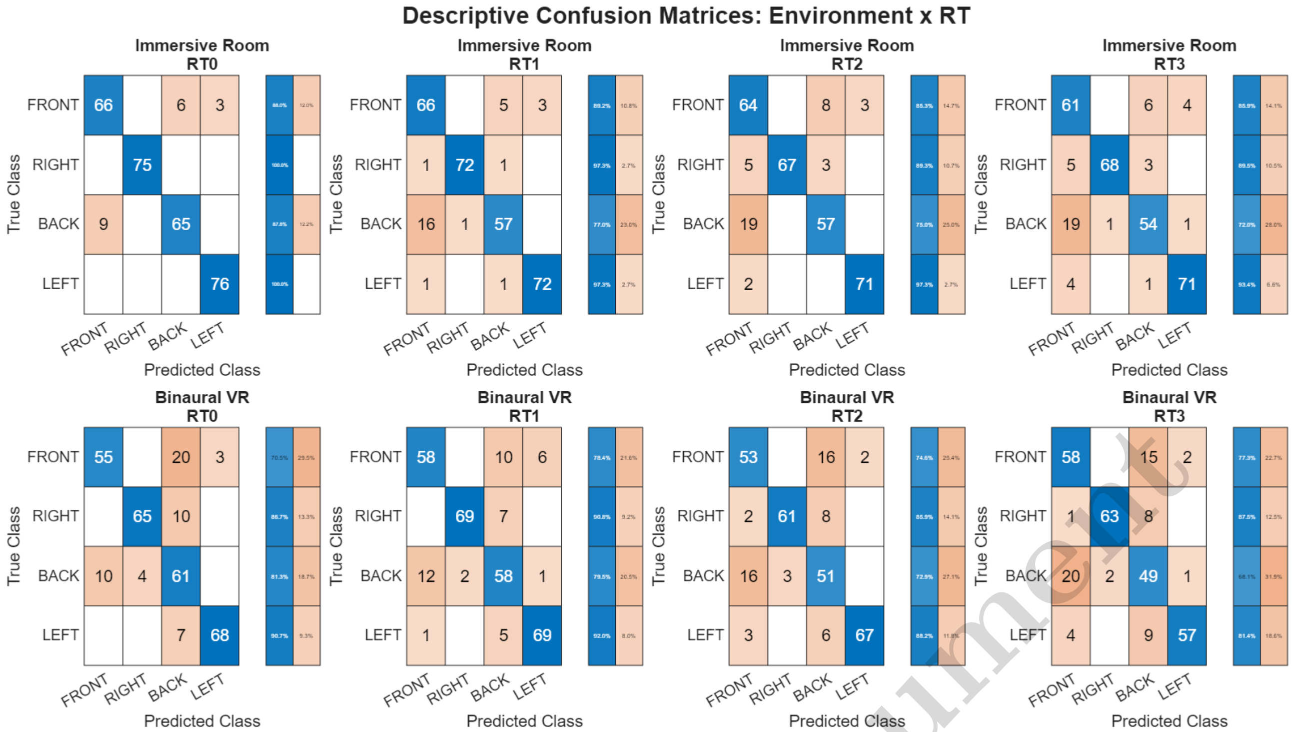


Figure 4: Confusion matrices aggregated across participants for the immersive room and binaural VR conditions.

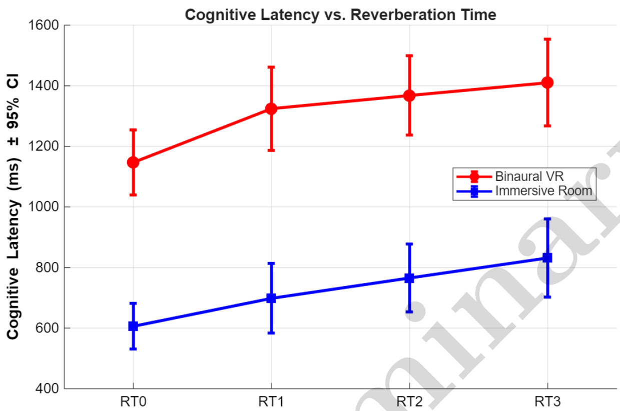


Figure 5: Mean latency-compensated response time as a function of reverberation condition for the immersive room and binaural VR.

based immersive reproduction and head-tracked binaural virtualization when simulating reverberant spaces.

5. Conclusions

This work has explored the influence of reverberation on horizontal sound source localization by comparing a 16.2.10 audiovisual immersive room with a head-tracked binaural VR system. The evaluation was conducted under four simulated acoustic conditions, ranging from quasi-anechoic to highly reverberant, derived from a reverberation chamber. The results in this work support the use of audiovisual immersive rooms as an effective platform for spatial audio reproduction under reverberant conditions. Across all reverberation conditions, the immersive room consistently achieved higher localization accuracy and shorter response latencies than the evaluated binaural VR implementation. Although both reproduction systems exhibited a degradation pattern as reverberation increased, the immersive room maintained a stable performance advantage of approximately 5–7 percentage points under reverberant conditions, reaching a maximum difference of 11.8 percentage points in the anechoic baseline, alongside substantially lower response times. The findings highlight the potential of loudspeaker-array-based immersive environments for spatial hearing experiments and perceptually demanding virtual acoustic applications. Future work will investigate how different immersive-room rendering strategies and room configurations influence localization performance under increasingly complex acoustic scenarios.

Acknowledgments

The authors would first like to express their sincere gratitude to all the volunteers who participated in the perceptual experiments for their time, effort, and dedication. Additionally, the authors extend their profound gratitude to d&b audiotechnik, and specifically to Juanma de Casas (Senior Consultant, Global Education and Application Support, d&b audiotechnik GmbH & Co. KG), for his invaluable support in coordinating and organizing the shipment of the entire electroacoustic equipment from Germany to our laboratory. The authors would also like to thank Gerard Díez Ruiz De Los Paños for his assistance in preparing the experimental setup.

References

[1] J. Blauert, *Spatial Hearing: The Psychophysics of Human Sound Localization*. Cambridge, MA: MIT Press, 1997.

[2] M. Tomasetti and L. Turchet, "Latency of spatial audio plugins: A comparative study," *IEEE Transactions on Human-Machine Systems*, vol. 53, no. 3, pp. 501–511, 2023.

[3] C. Mendonça, G. Campos, P. Dias, and J. A. Santos, "On the improvement of auditory accuracy with non-individualized hrtf-based sounds," *Journal of the Audio Engineering Society*, 2012.

[4] P. Majdak, M. J. Goupell, and B. Laback, "3-d localization of virtual sound sources: Effects of visual environment, pointing method, and training," *Attention, Perception, & Psychophysics*, vol. 72, no. 2, pp. 454–469, 2010.

[5] S. Fargeot, O. Derrien, G. Parseihian, M. Aramaki, and R. Kronland-Martinet, "Perceptual evaluation of an ambisonic auralization system of measured 3d acoustics," *Acta Acustica*, vol. 7, p. 56, 2023.

[6] A. Guastamacchia, R. Rosso, and F. Riente, "Real and virtual lecture rooms: Objective and subjective validation of a 16-speaker-based virtual reality system," *Acoustics*, vol. 6, pp. 938–960, 2024.

[7] F. Zotter and M. Frank, *Ambisonics: A Practical 3D Audio Theory for Recording, Studio Production, Sound Reinforcement, and Virtual Reality*. Springer Nature, 2019.

[8] C. Scheer, F. Lau, C. Sander, S. Weinzierl, and J. M. Arend, "The influence of visual cues on the sound externalization of virtual loudspeaker setups," *Journal of the Audio Engineering Society*, vol. 74, no. 1/2, pp. 5–19, 2026.

[9] A. Ahrens, K. D. Lund, M. Marschall, and T. Dau, "Sound source localization with varying amount of visual information in virtual reality," *PLoS ONE*, vol. 14, no. 3, e0214603, 2019.

[10] S. Fichna et al., "Perception of simulated moving sound sources in reverberant virtual audiovisual environments," in *Proceedings of Forum Acusticum 2025*, Málaga, Spain, 2025.

[11] A. J. Kolarik, B. C. J. Moore, P. Zahorik, S. Cirstea, and S. Pardhan, "Auditory distance perception in humans: A review of cues, development, neuronal bases, and effects of sensory loss," *Attention, Perception, & Psychophysics*, vol. 78, pp. 373–395, 2016.

[12] J. Catic, S. Santurette, and T. Dau, "The role of reverberation-related binaural cues in the externalization of speech," *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 1154–1167, 2015.

[13] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," *The Journal of the Acoustical Society of America*, vol. 111, no. 5, pp. 2110–2117, 2002.

[14] T. Lübeck et al., "Perceptual evaluation of different methods for binaural rendering of recordings with various microphone arrays," *Journal of the Audio Engineering Society*, vol. 74, no. 5, pp. 256–269, 2026. DOI: `10.17743/jaes.2022.0259`.

[15] L. McCormack and A. Politis, "Sparta & compass: Real-time implementations of linear and parametric spatial audio reproduction and processing methods," in *Proceedings of the AES International Conference on Immersive and Interactive Audio*, 2019.

[16] Institute of Electronic Music and Acoustics (IEM), *IEM Plugin Suite*, `https://plugins.iem.at/`, University of Music and Performing Arts Graz, 2024.

[17] P. Gutierrez-Parera, F. Arrebola, M. Cuevas-Rodriguez, D. Gonzalez-Toledo, L. Molina-Tanco, and A. Reyes-Lecuona, "Exploring simple methods for room absorption estimation in early reflections for binaural audio rendering," in *Proceedings of the 11th Convention of the European Acoustics Association (Forum Acusticum)*, Málaga, Spain, Jun. 2025.