AUDITORY DISCRIMINATION OF EARLY REFLECTIONS IN VIRTUAL ROOMS

Junting Chen
School of Computing, Science & Engineering
University of Salford
Manchester, UK
J.Chen14@edu.salford.ac.uk

Duncan Williams
School of Computing, Science & Engineering
University of Salford
Manchester, UK
D.A.H.Williams@edu.salford.ac.uk

Bruno Fazenda
School of Computing, Science & Engineering
University of Salford
Manchester, UK
B.M.Fazenda@edu.salford.ac.uk

ABSTRACT

This study investigates the perceptual sensitivity to early reflection changes across different spatial directions in a virtual reality (VR) environment. Using an ABX discrimination paradigm, participants evaluated speech stimuli convolved with third-order Ambisonic room impulse responses under three position reversal (Left–Right, Front–Back, and Floor–Ceiling) and three reverberation conditions (RT60 = 1.0 s, 0.6 s, and 0.2 s). Binomial tests revealed that participants consistently detected early reflection differences in the Left–Right reversal, while discrimination performance in the other two directions remained at or near chance. This result can be explained by the higher acuity and lower localisation blur found for the human auditory system. A two-way ANOVA confirmed a significant main effect of spatial position ($p = 0.00685$, $\eta^2 = 0.1605$), with no significant effect of reverberation or interaction. The analysis of the binaural room impulse responses showed wave forms and Direct-Reverberant-Ratio differences in the Left–Right reversal position, aligning with perceptual results. However, no definitive causal link between DRR variations and perceptual outcomes can yet be established.

1. INTRODUCTION

As sound travels through a room, it reaches the listener not only directly but also after reflecting off one or more surfaces, giving rise to early reflections and reverberation over time. Early reflections provide cues that help listeners perceive the size and geometry of their surroundings, while reverberation influences how sound decays and merges within the acoustic space.

In virtual reality, room acoustics can be simulated using either real-time or pre-calculated approaches. Real-time methods perform all acoustic simulation and spatialization during playback but require significant simplifications to meet latency constraints (typically <100 ms). In contrast, pre-calculated methods allow for more accurate modelling by computing room impulse responses (RIRs) in advance, often across a spatial grid of receiver positions, and storing them in spatial formats such as Ambisonics for later binaural decoding. Two primary strategies are used for simulating room acoustics: geometrical acoustics and wave-based models. Geometrical methods, which approximate sound as rays reflecting off surfaces, are computationally efficient and well-suited for large rooms, but may fail to capture diffraction and interference effects in smaller spaces. Wave-based methods, including the Boundary Element Method (BEM) and Finite Element Method (FEM), offer greater accuracy for low-frequency and small-room simulations, though at significantly higher computational cost. Regardless of the method, the outcome is typically RIR and derived parameters such as reverberation time and clarity, which are convolved with anechoic signals and spatialized for reproduction. However, simulation accuracy can be affected by input data quality, algorithmic limitations, and uncertainties in boundary conditions or model geometry. These factors must be considered when selecting an appropriate modelling approach for VR applications.

Among the various components of room acoustic modelling, early reflections (ERs) have received particular attention due to their perceptual relevance. Changes in early reflections have been shown to produce perceptible variations in timbre [1]. They also affect the perceived distance and clarity of sound, thereby enhancing the realism and intelligibility of auditory scenes, particularly in communication settings [2]. Moreover, early reflections contribute to accurate sound source localization and play a key role in the perception of sound directionality in virtual space [3]. Reflections occurring within the first 80 milliseconds after the direct sound are especially important for reinforcing the realism of perceived source locations [4]. Simulations that incorporate both the direct sound path and first-order reflections have been shown to improve the overall perceptual quality of virtual environments [5]. Accordingly, the effective management of early reflections is essential for creating believable and immersive virtual acoustic environments [6].

Although the perceptual properties of early reflections have been well established and widely applied in previous studies, much of this work has not been conducted within virtual reality environments. Modern VR audio systems integrate multiple techniques, including room acoustic modelling, auralization, and spatial rendering, into immersive platforms that often feature dynamic head tracking and real-time three-dimensional interactivity. Within such complex systems and specific play-back conditions, a critical and practical question arises: when applied in a full VR spatial audio rendering context, are modelling refinements—such as changes in the early reflection sound, still perceptible to the listener?

This study investigates the perceptual discriminability of changes in early reflections within a virtual reality environment, where variations in acoustic modelling are introduced by manipulating the source–receiver configuration and the surface absorption coefficients of the room.

This work investigates two research questions:

(1) Can listeners detect perceptual differences resulting from changes in early reflections in a VR spatial audio environment?

(2) Does reverberation time influence the ability to discriminate against these differences?

2. EXPERIMENT

The aim of the experiment was to examine whether variations in early reflections, controlled through changes in room acoustic modelling, could be perceptually discriminated by listeners in a virtual reality (VR) environment. Specifically, the study investigated how spatial manipulations along different axes—while keeping the visual scene and relative source–receiver distance constant—affected the detectability of early reflection differences. Additionally, the experiment explored whether changes in overall reverberation time would influence the perceptual discrimination of these early reflections.

2.1. Apparatus

The stereoscopic headset used in this study was the Meta Quest 2, operated at a 60 Hz refresh rate to ensure visual comfort and minimize the risk of motion sickness. No frame drops or discomfort were reported by participants. The experiment was run on a Windows 10 laptop equipped with an NVIDIA RTX 3070 graphics card, using Unity version 2022.3.11f1 for visual rendering and experimental control. The audio was played back through a pair of Sennheiser HD650 open-back headphones, which were set to the same listening level for each subject.

2.2. Stimuli Generation

Acoustic modelling was conducted using CATT-Acoustic v9. A virtual rectangular room was defined, and four source–receiver configurations were created: one reference position and three comparison positions, as illustrated in **Figure 1**. In the reference setup, the source was placed 1 meter from the front wall, 1 meter from the right wall, and 0.8 meters above the floor. Across all configurations, the source and receiver were spaced 1.2 meters apart, with aligned height and azimuth to eliminate variations in elevation and horizontal angular offset. This configuration was designed to isolate changes in early reflections along a single spatial dimension in each comparison, relative to the reference position. Placing the reference position near the corner of the room also maximized spatial separation between conditions, thereby enhancing the acoustic contrast across test scenarios.

An anechoic female speech clip was used as the stimulus, selected based on preliminary self-assessment indicating that speech provides greater sensitivity to acoustic variations under the current experimental conditions than other sources such as music. This clip was convolved in MATLAB with third order Ambisonic room impulse responses (RIRs) simulated in CATT-Acoustic. The resulting Ambisonic signals were then imported into Wwise and decoded binaurally using the Google Resonance Spatializer, enabling head-tracked spatial audio presentation within the VR environment.

Ambisonics is a system for capturing and reproducing three-dimensional spatial audio using spherical harmonics to describe the sound field. A key concept in this system is the Ambisonic order, which determines the spatial resolution of the encoded audio. The first order system [7] (commonly referred to as B-format) uses four channels: one omnidirectional (W) and three figure-of-eight microphones aligned with the X, Y, and Z axes to encode horizontal and vertical directional cues [8], respectively. These represent the zeroth-order pressure and first order particle velocity components [9]. Higher-order Ambisonics (HOA) extend this principle by including additional channels that capture finer directional information, with each order adding more spherical harmonic components, meaning higher orders include all lower-order components [10]. For example, third order Ambisonics requires 16 channels in total, providing improved angular resolution, more accurate wavefront reconstruction, and a larger perceptual "sweet spot" during playback [11]. In particular, the precision of direct sound reproduction has been shown to benefit from increased Ambisonic order [12].

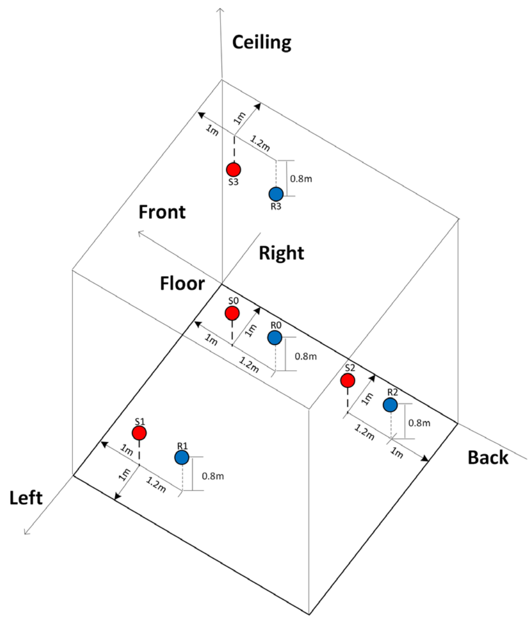

Figure 1: *Spatial configurations of the source–receiver pairs in the virtual room (7 m × 5 m × 4 m). S denotes the sound source and R denotes the receiver. Red and blue markers are used to distinguish sound sources (S) and receivers (R), respectively. S_0–R_0 represents the reference position, while S_1–R_1 (Left–Right), S_2–R_2 (Front–Back), and S_3–R_3 (Floor–Ceiling) denote spatial comparisons along a single axis. In all configurations, the source–receiver distance was fixed at 1.2 m with no elevation or azimuthal offset.*

Third order Ambisonics was chosen as a balance between spatial fidelity and practical implementation constraints. While Ambisonic orders beyond third can theoretically offer even higher accuracy, they impose substantial demands on both the rendering environment and decoding systems. Among the available spatialization toolkits, only Wwise supports Ambisonics beyond first order, whereas other mainstream spatializers—such as Oculus Spatializer, Steam Audio, and Unity's native audio engine—typically provide support only for first-order encoding. For this reason, Wwise was selected and integrated with Unity to build the VR environment. Third-order Ambisonics thus represented the highest order that could be robustly supported within our experimental.

2.3. Methodology

Participants were instructed to complete an ABX discrimination task, in which they were asked to determine whether stimulus X was perceptually identical to stimulus A or stimulus B. The ABX test is structured as a series of Bernoulli trials— each representing a binary outcome where a correct response (correctly identifying whether X matches A or B) is considered a ‘success,’ and an incorrect response is treated as a ‘failure.’ Repeating such independent trials under the same probability model results in a binomial distribution of outcomes. Although prior studies such as R.E.Greenaway [13] emphasize the statistical value of using a larger number of assessors over repeating trials with a smaller group, practical constraints often necessitate a trade-off [14]. In this study, each stimulus pair was repeated 10 times per participant per condition to ensure enough observations.

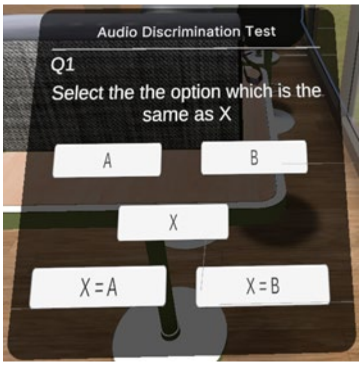

Figure 2: *Screenshot of the in-VR questionnaire panel used during the ABX test. Participants were presented with looped playback controls for samples A, B, and X, followed by two response buttons to identify which of A or B matched X.*

After completing an initial calibration and instruction phase, participants practiced using the interface, which featured a virtual room and a floating questionnaire panel with five interactive buttons (see **Figure 2**). The top three buttons (A, B, and X) triggered looped playback of the respective audio stimuli, and participants could freely toggle between them. Once ready, they used the two response buttons (“X = A” or “X = B”) to submit their choice. To minimize any order bias, the sample presentation sequence was randomized in each trial.

Table 1: *Experimental conditions across position comparisons and reverb levels.*

Position Comparisons	Reverb	Number of trials
Left – Right	1s	3 x 3x 10 = 90
Front – Back	0.6s	
Floor – Ceiling	0.2s	

Following the familiarization stage, participants completed the experiment independently in a quiet, private space. The full session comprised 90 trials per participant, derived from 3 spatial conditions (Left–Right, Front–Back, Floor–Ceiling) × 3 reverberation settings (RT60 ≈ 1.0 s, 0.6 s, 0.2 s) × 10 repetitions (see **Table 1**). Upon completion, the system automatically submitted all responses via an integrated Google Form for further statistical analysis.

2.4. Subjects

Seven participants aged between 24 and 35, took part in this stage of the study. All reported normal hearing and had no prior training in audio perception tasks. Each participant provided informed consent and was instructed thoroughly before beginning the test. While the number of participants was relatively small (N = 7), each participant completed 10 independent and randomized trials per condition, resulting in 70 observations per test condition. The trial design ensured statistical independence and minimized order effects. Furthermore, the statistical power analysis reported in the following section is an indicator that sufficient sampling has occurred.

3. RESULTS

This section presents the statistical analysis of participants' performance in the ABX discrimination task. The goal was to determine whether changes in early reflections across different spatial configurations and reverberation conditions could be reliably perceived. Two complementary analyses were conducted: a binomial test to assess whether response accuracy exceeded chance levels under each condition, followed by a two-way ANOVA to evaluate the main and interaction effects of spatial position and reverberation time on perceptual accuracy.

3.1. Binominal Test Analysis

To assess whether participants were able to discriminate early reflection differences above chance level, a binomial test was conducted on responses collected from seven participants. Each participant completed 90 ABX trials across nine experimental conditions (3 spatial configurations × 3 reverberation settings), with 10 repetitions per condition. The randomized trial order and independence of ABX responses ensured the validity of the binomial analysis.

Table 2: *Binomial Test Outcomes by Position Comparison and Reverberation.*

Position / **Reverb**	Front–Back	p-value	Left–Right	p-value	Floor–Ceiling	p-value
1s	39/70	0.20	51/70	**$8.3e^{-5}$**	38/70	0.28
0.6s	44/70	**0.02**	47/70	**0.003**	42/70	0.06
0.2s	35/70	0.55	53/70	**$9.6e^{-6}$**	40/70	0.14

Binomial test results revealed marked differences in participants' ability to discriminate early reflection changes across spatial directions and reverberation times (RT60), as summarized in **Table 2**.

For the Left–Right reversal, performance was consistently above chance across all reverberation conditions. Participants achieved 51/70 correct at RT60 = 1.0 s ($p = 8.3e^{-5}$), 47/70 at 0.6 s ($p = 0.003$), and 53/70 at 0.2 s ($p = 9.6e^{-6}$), indicating robust perceptual sensitivity to horizontal directional changes in early reflections.

In the Front–Back reversal, only the medium reverberation level (0.6 s) resulted in a statistically significant outcome (44/70, $p = 0.02$). The other two conditions (RT60 = 1.0 s and 0.2 s) did not reach significance, with 39/70 ($p = 0.20$) and 35/70 ($p = 0.55$), respectively, suggesting that Front–Back reversal were less reliably perceived.

For the Floor–Ceiling reversal, no significant effects were observed. Accuracy remained near chance level across all three reverberation settings: 38/70 ($p = 0.28$) at 1.0 s, 42/70 ($p = 0.06$) at 0.6 s, and 40/70 ($p = 0.14$) at 0.2 s. These results indicate that vertical shifts in early reflections were the least perceptually salient.

In summary, participants exhibited the highest sensitivity to Left–Right spatial changes, modest sensitivity in Front–Back under moderate reverberation, and minimal perceptual response to Floor–Ceiling variations.

3.2. Anova Analysis

To further assess the influence of spatial configuration and reverberation on early reflection discrimination, a two-way ANOVA was conducted with participants' ABX accuracy (number of correct responses) as the dependent variable. Assumption checks were performed prior to the analysis. Assumption checks were performed prior to the analysis. The Shapiro–Wilk test conducted on the overall dataset yielded a p-value of 0.0057, suggesting a deviation from normality at the aggregate level. However, when examined at the group level (i.e., across all 9 condition combinations), all Shapiro–Wilk p-values exceeded 0.05, indicating that the data within each condition could be considered normally distributed. This conclusion was further supported by visual inspection of the Q–Q plot, which showed approximate linear alignment of the data points. Levene's test for homogeneity of variance also yielded a non-significant result ($p = 0.741$), confirming equality of variances across groups. Based on these combined results, the data were deemed suitable for parametric testing using a two-way ANOVA.

Table 3: *Two-Way ANOVA Summary Table.*

Factors	p-value	Effect size (Eta^2)	Power
Position	**0.00685**	0.1605	1.000
Reverb	0.87556	0.0039	0.2294
Interaction: Position × Reverb	0.56843	0.0434	0.9803

The analysis revealed a significant main effect of Position ($p = 0.00685$), with a large effect size ($\eta^2 = 0.1605$), indicating that the spatial configuration of source–receiver pairs significantly influenced participants' ability to detect changes in early reflections. In contrast, the reverberation factor, representing reverberation differences across RT60 values, did not yield a statistically significant effect ($p = 0.87556$, $\eta^2 = 0.0039$). Similarly, the Position × reverberation interaction was non-significant ($p = 0.56843$, $\eta^2 = 0.0434$), suggesting no meaningful interaction between spatial orientation and reverberation level on perceptual accuracy (see **Table 3**). However, the statistical power for the reverberation factor was relatively low (power = 0.2294), which leaves open the possibility of a Type II error—failing to detect an existing effect. Although the results do not support a robust influence of reverberation under current conditions, this low power suggests that subtle perceptual effects may have gone undetected and could warrant further exploration under refined experimental conditions.

Table 4: *Tukey HSD Test Results for Position Comparisons.*

Position	p-value
(Left-Right) x (Front-Back)	**0.013**
(Front-Back) x (Floor-Ceiling)	0.982
(Left-Right) x (Floor-Ceiling)	**0.020**

To further investigate the role of sensitivity to spatial orientation, post hoc comparisons using the Tukey HSD test were conducted. Results revealed that Left–Right reversal led to significantly higher discrimination performance compared to both Front–Back ($p = 0.013$) and Floor–Ceiling ($p = 0.020$) reversals. However, no significant difference was found between Front–Back and Floor–Ceiling ($p = 0.982$). These findings support the interpretation that early reflection differences along the horizontal (Left–Right) axis are more perceptually salient than those along the depth or vertical axes (see **Table 4**).

Visualizations using notched boxplots further support the statistical findings. As shown in **Figure 3**, participants exhibited consistently higher correct response rates under the Left–Right reversal group across all reverberation times, with elevated medians and relatively narrow interquartile ranges. In contrast, the Front–Back and the Floor–Ceiling reversals showed lower medians and greater variability, particularly at longer reverberation times (RT60 = 1.0 s). These patterns reinforce the observation that spatial orientation, rather than reverberation level alone, played a more decisive role in perceptual discrimination performance.

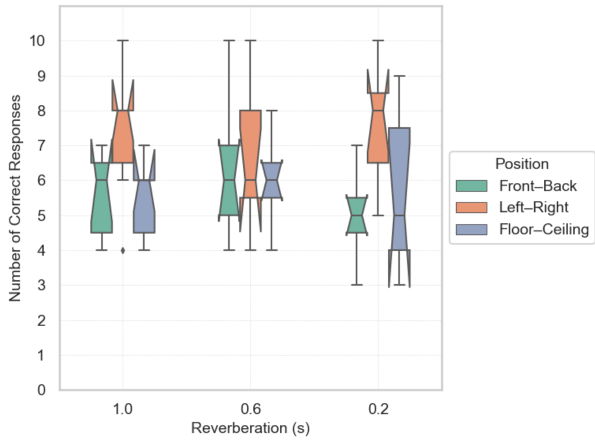

Figure 3: *Notched boxplot showing the number of correct responses grouped by (RT60 time in seconds), with Position Comparison distinguished by colour.*

4. DISCUSSION

The binomial tests provided an initial indication of perceptual sensitivity by evaluating whether participants' responses exceeded chance levels under each experimental condition. Results showed that in the Left–Right reversal, participants consistently achieved statistically significant detection rates across all three reverberation conditions (RT60 = 1.0 s, 0.6 s, and 0.2 s). This suggests that early reflection differences along the horizontal axis were perceptually salient, regardless of reverberation level. For the Front–Back reversal, significance was only found at the medium reverberation level (RT60 = 0.6 s), indicating some sensitivity to changes in this direction, though less robust. In the Floor–Ceiling reversal, none of the reverberation levels resulted in performance significantly above chance, implying that vertical variations in early reflections were largely imperceptible to participants under the current test setup.

Complementing these results, the two-way ANOVA identified a significant main effect of Position ($p = 0.00685$) with a large effect size ($\eta^2 = 0.1605$), confirming that the spatial axis along which the source and receiver were displaced had a meaningful impact on detection accuracy. However, no significant effect was found for reverberation ($p = 0.87556$), nor for the interaction between Position and reverberation ($p = 0.56843$), indicating that reverberation time and its interplay with spatial orientation did not significantly influence performance. This conclusion is supported by Tukey HSD post hoc analysis, which revealed that the Left–Right reversal differed significantly from both Front–Back and Floor–Ceiling, while no significant difference was observed between the latter two.

Together, these findings emphasize the perceptual dominance of early reflection changes in the horizontal (Left–Right) direction and suggest that participants were relatively insensitive to early reflection variations in the median and the frontal plane. This result can be explained by the higher acuity and lower localisation blur found for the human auditory system [15]. While reverberation level alone did not systematically affect performance, mid-level reverberation (RT60 = 0.6 s) may have introduced greater perceptual contrast in some conditions, a hypothesis that warrants further investigation in future studies.

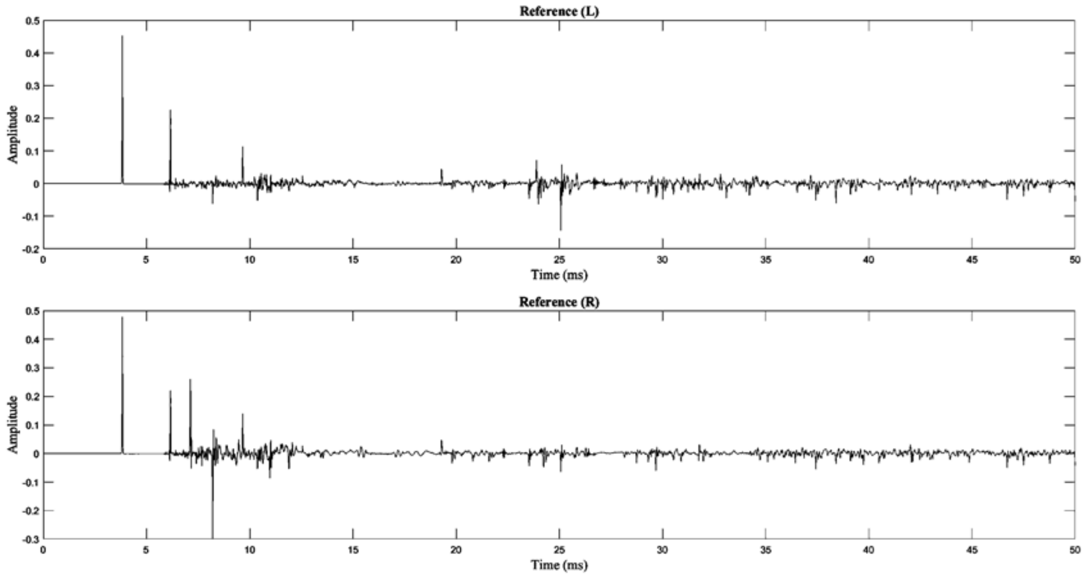

Figure 4: *Binaural output waveform of the reference source–receiver configuration. The signals shown are the left and right channel responses rendered from third-order Ambisonic room impulse responses (RIRs) simulated in CATT-Acoustic and imported directly into the audio system without convolution. This represents the left and right reference position channels.*

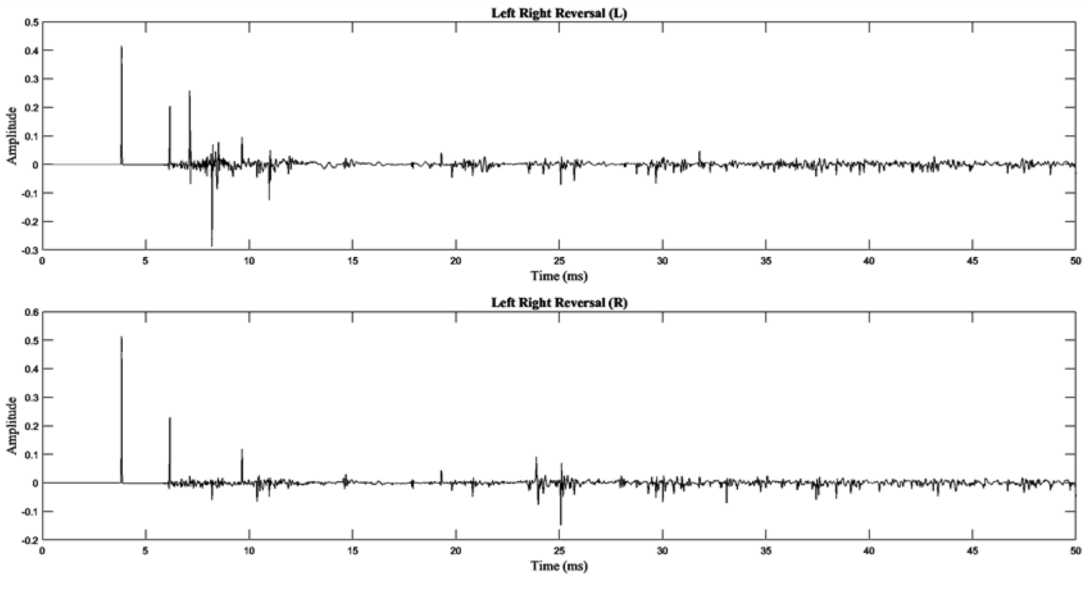

Figure 5: *Binaural output waveform of the left-right reversal configuration. The signals shown are the left and right channel responses rendered from third-order Ambisonic room impulse responses (RIRs) simulated in CATT-Acoustic and imported directly into the audio system without convolution. This represents the Left-Right reversal position channels.*

Further insight is gained through analysis of the simulated room impulse responses (RIRs). Judging from the binaural room impulse response (RIR) waveforms derived from the test audio system, the Left–Right reversal (see **Figure 5**) reveals expected changes relative to the reference configuration (see **Figure 4**). The reflection components arriving laterally have been inverted, whilst medial and frontal plane reflection patterns have been maintained.

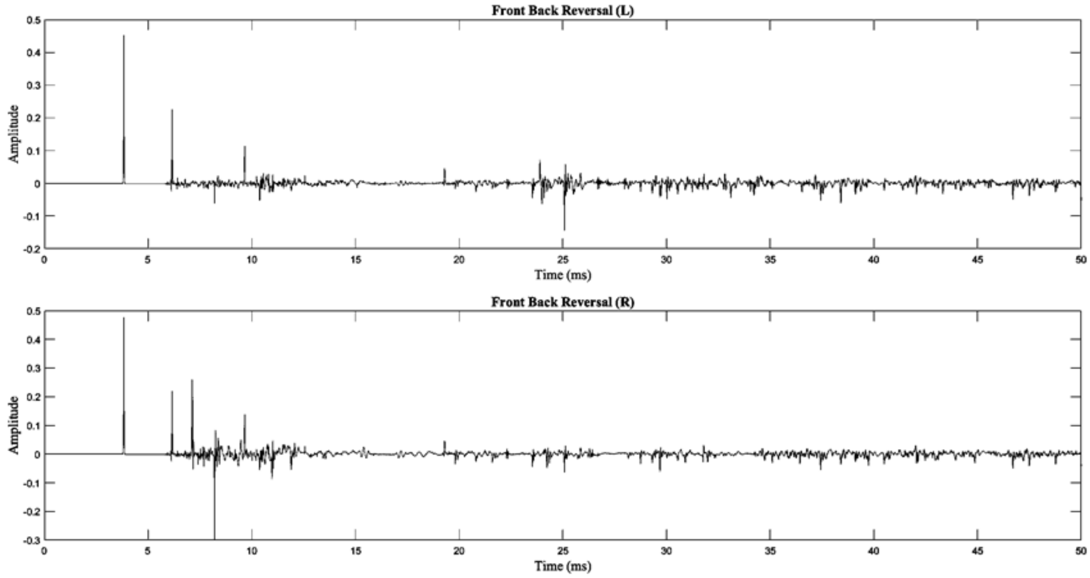

Figure 6: *Binaural output waveform of the Front-Back reversal configuration. The signals shown are the left and right channel responses rendered from third-order Ambisonic room impulse responses (RIRs) simulated in CATT-Acoustic and imported directly into the audio system without convolution. This represents the Front-Back reversal position channels.*

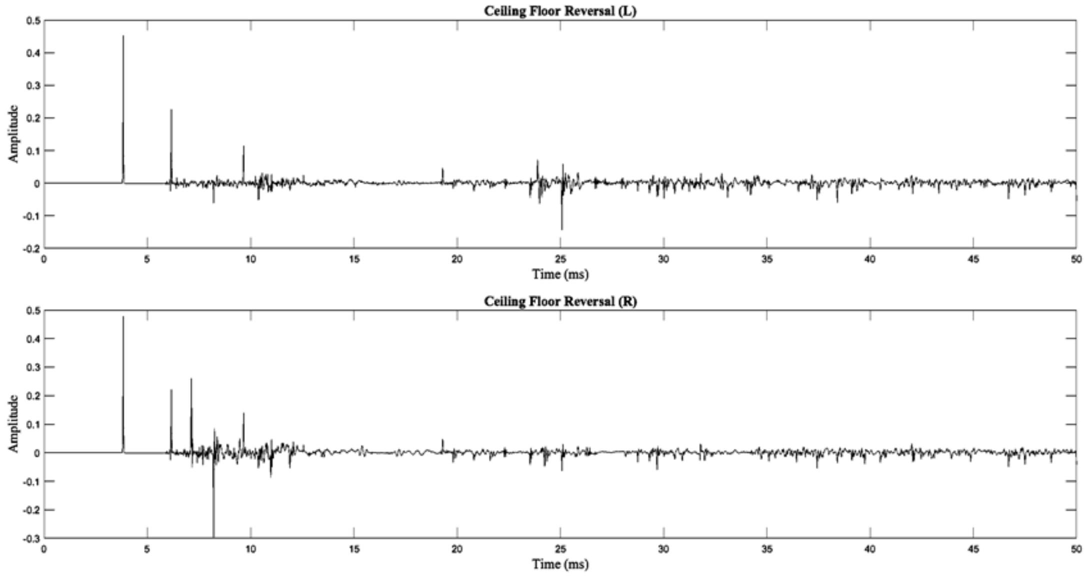

Figure 7: Binaural output waveform of the Floor-Ceiling reversal configuration. The signals shown are the left and right channel responses rendered from third-order Ambisonic room impulse responses (RIRs) simulated in CATT-Acoustic and imported directly into the audio system without convolution. This represents the Floor-Ceiling reversal position channels.

In contrast, the Front–Back and Floor–Ceiling reversals (see **Figure 6** and **Figure 7**) exhibit only minor amplitude differences between corresponding channels, with no substantial shifts in arrival times or interaural structure. These subtle acoustic differences help explain why participants were generally unable to discriminate these configurations in the listening tests. Overall, the wave-form data supports the observed perceptual sensitivity to early reflections in the horizontal plane.

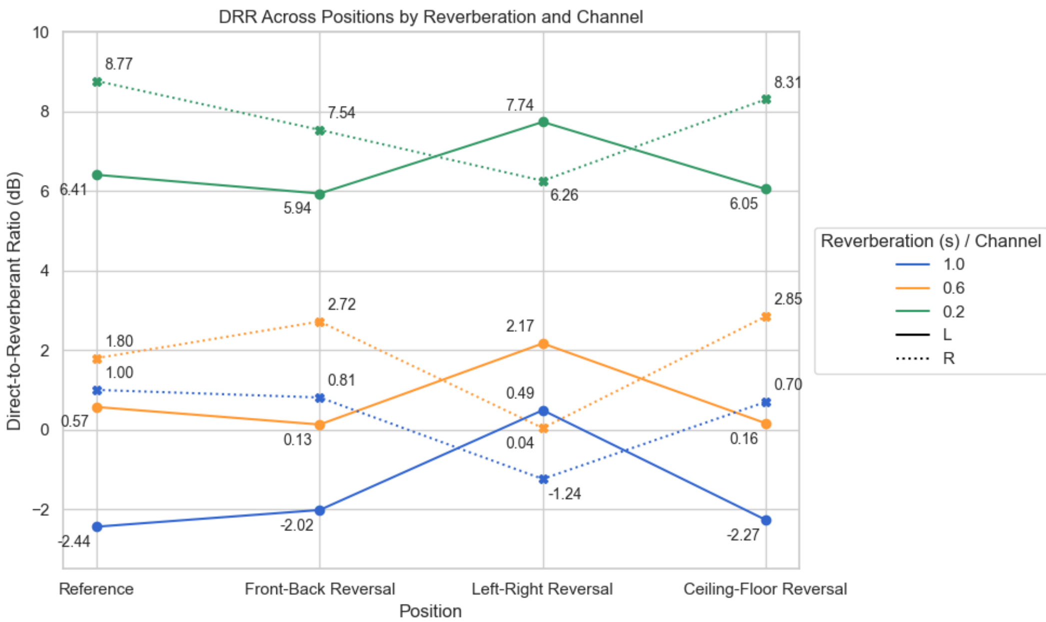

Figure 8: *Line plot illustrating the Direct-to-Reverberant Ratio (DRR) across four spatial configurations (Reference, Front–Back Reversal, Left–Right Reversal, Ceiling–Floor Reversal), for three reverberation conditions (RT60: 1.0 s, 0.6 s, 0.2 s). Solid lines represent the left (L) binaural channel, and dotted lines represent the right (R) channel. DRR is expressed in decibels (dB).*

The analysis includes the Direct-to-Reverberant Ratio (DRR), a commonly used acoustic metric that quantifies the energy relationship between early reflections and late reverberation. In this study, DRR was computed using the implementation in the IoSR MATLAB toolbox [16], based on the following equation [17]:

$$DRR = 10log_{10}\left(\frac{\sum_{n=T_0-C}^{n=T_0+C} X^2(n)}{\sum_{n=T_0+C+1}^{N} X^2(n)}\right) \quad (1)$$

Where X(n) is the approximated integral of the impulse response, T0 is the time of the direct impulse, which was estimated automatically, and C is a symmetric 2.5 ms window around T0 used to separate the early energy from the late reverberant tail.

It is worth noting that the Left–Right reversal consistently produced the most pronounced DRR deviations across all three reverberation settings (see **Figure 8**). In contrast to the relatively stable trends observed in the Front–Back and Ceiling–Floor reversals, the Left–Right reversal exhibits sharp rises and drops, particularly when compared against the reference position.

While these variations do not necessarily establish a direct causal relationship with perceptual outcomes, they offer a possible acoustic explanation for the observed discriminability pattern. In the Left–Right reversal, lateral displacements of the source–receiver pair may have concentrated early reflection energy toward one ear, resulting in greater interaural energy differences. This asymmetry is reflected in the larger per-channel DRR differences. In contrast, the Front–Back and Ceiling–Floor reversals tend to maintain interaural balance, yielding comparatively smaller DRR shifts.

Therefore, although the observed DRR asymmetries correspond with the direction-sensitive perceptual results, further investigation is needed to clarify whether and how these energy distributions directly influence early reflection discrimination in binaural contexts.

5. CONCLUSION

This study examined the perceptual discriminability of early reflection changes in a virtual reality (VR) environment, focusing on spatial variations introduced along three orthogonal axes—Left–Right, Front–Back, and Floor–Ceiling—under multiple reverberation conditions. Participants engaged in an ABX discrimination task to evaluate whether audio stimuli, derived from third-order Ambisonic room impulse responses convolved with speech signals, could be differentiated based on early reflection cues alone.

Consistent findings from both binomial tests and a two-way ANOVA confirmed that early reflection changes along the Left–Right (horizontal) axis are perceptually salient, yielding significantly above-chance accuracy across all reverberation settings. In contrast, changes along the Front–Back and Floor–Ceiling axes failed to elicit reliable discrimination. Statistical analysis revealed a significant main effect of Position with a large effect size ($\eta^2 = 0.1605$) and a strong statistical power (1.000), while neither the reverberation factor nor its interaction with Position reached significance—suggesting that reverberation time had minimal perceptual influence in this context. These results are to be expected given the higher acuity in source localisation for horizontal plane in humans. However, results pertaining to sensitivity to changes in the early reflection sound field had not been demonstrated in literature hitherto.

Signal-level analyses of the binaural impulse responses further support these behavioural outcomes. The Left–Right reversal exhibited pronounced changes in both early reflection structure and Direct-to-Reverberant ratio (DRR), these acoustical features were less distinct in the other two directions, aligning with participants' perceptual insensitivity. Together, these results indicate that lateral asymmetries in early reflections are more perceptually relevant than vertical or depth-based variations in immersive environments.

Although the current study approximated the room shape, sound source position, listener location, and source–listener distance in the VR visuals to match the acoustic model, these elements were not perfectly aligned. As a result, audiovisual consistency could not be examined as a controlled independent variable. Nonetheless, this partial mismatch still introduces implicit cross-modal incongruence, underscoring the need to explore its perceptual consequences more systematically. If the visual space were fully matched to the acoustic model, future experiments could directly investigate the perceptual effects of audiovisual coherence.

Prior study has established that cross-modal congruency plays a critical role in enhancing presence in virtual environments [18-20]. Reverberation has been shown to reinforce presence, externalization, and naturalness [21, 22]. However, the specific perceptual role of early reflections within this context remains the subject of further work.

Future research should prioritize examining how specific acoustic characteristics of early reflections—such as their timing, directionality, and relative energy—affect spatial perception in VR contexts. In addition, studies should explore the relationship between reverberation parameters (e.g, DRR, RT60) and broader perceptual constructs such as presence, realism, and plausibility, to better understand how acoustic design influences user experience in immersive environments.

6. ACKNOWLEDGMENTS

The authors would like to thank all participants for their time and commitment during the perceptual tests. This work was supported by School of Computing, Science & Engineering. The authors also wish to acknowledge the developers of the open-source toolboxes and software platforms that made this work possible.

7. REFERENCES

[1] D. R. Begault, B. U. McClain, and M. R. Anderson, "Early reflection thresholds for virtual sound sources," in Proc. 2001 Int. Workshop on Spatial Media, 2001.

[2] Y. Kanada, "Multi-context voice communication in a SIP/SIMPLE-based shared virtual sound room with early reflections," in Proceedings of the international workshop on Network and operating systems support for digital audio and video, 2005, pp. 45-50.

[3] H. Steffens, S. van de Par, and S. D. Ewert, "The role of early and late reflections on perception of source orientation," The Journal of the Acoustical Society of America, vol. 149, no. 4, pp. 2255-2269, 2021.

[4] B. Kapralos, "Auditory perception and virtual environments," York University, Canada, 2003.

[5] N. Tsingos, E. Gallo, and G. Drettakis, "Perceptual audio rendering of complex virtual environments," ACM Transactions on Graphics (TOG), vol. 23, no. 3, pp. 249-258, 2004.

[6] E. M. Wenzel, "What perception implies about implementation of interactive virtual acoustic environments," in Audio Engineering Society Convention 101, 1996: Audio Engineering Society.

[7] D. G. Malham and A. Myatt, "3-D sound spatialization using ambisonic techniques," Computer music journal, vol. 19, no. 4, pp. 58-70, 1995.

[8] R. K. Furness, "Ambisonics-an overview," in Audio Engineering Society Conference: 8th International Conference: The Sound of Audio, 1990: Audio Engineering Society.

[9] N. Epain, C. Jin, and F. Zotter, "Ambisonic decoding with constant angular spread," Acta Acustica united with Acustica, vol. 100, no. 5, pp. 928-936, 2014.

[10] J. G. Tylka and E. Y. Choueiri, "Comparison of techniques for binaural navigation of higher-order ambisonic soundfields," in 139th Audio Engineering Society International Convention, AES 2015, 2015.

[11] S. Bertet, J. Daniel, E. Parizet, and O. Warusfel, "Investigation on localisation accuracy for first and higher order ambisonics reproduced sound sources," Acta Acustica united with Acustica, vol. 99, no. 4, pp. 642-657, 2013.

[12] S. Moreau, J. Daniel, and S. Bertet, "3d sound field recording with higher order ambisonics–objective measurements and validation of a 4th order spherical microphone," in 120th Convention of the AES, 2006, pp. 20-23.

[13] R. E. Greenaway, "ABX discrimination task," in Discrimination Testing in Sensory Science: Elsevier, 2017, pp. 267-288.

[14] H. T. Lawless and H. Heymann, Sensory evaluation of food: principles and practices. Springer Science & Business Media, 2010.

[15] J. Blauert, Spatial hearing: the psychophysics of human sound localization. MIT press, 1997.

[16] IoSR MATLAB Toolbox. (2025). University of Surrey. Accessed: 3/31/2025. [Online]. Available: https://github.com/IoSR-Surrey/MatlabToolbox

[17] P. Zahorik, "Direct-to-reverberant energy ratio sensitivity," The Journal of the Acoustical Society of America, vol. 112, no. 5, pp. 2110-2117, 2002.

[18] P. Larsson, D. Västfjäll, P. Olsson, and M. Kleiner, "When what you hear is what you see: Presence and auditory-visual integration in virtual environments," in Proceedings of the 10th annual international workshop on presence, 2007, pp. 11-18.

[19] M. Brengman, K. Willems, and L. De Gauquier, "Customer engagement in multi-sensory virtual reality advertising: the effect of sound and scent congruence," Frontiers in Psychology, vol. 13, p. 747456, 2022.

[20] O. Rummukainen, "Reproducing reality: Perception and quality in immersive audiovisual environments," 2016.

[21] P. Larsson, D. Västfjäll, and M. Kleiner, "Effects of auditory information consistency and room acoustic cues on presence in virtual environments," Acoustical Science and Technology, vol. 29, no. 2, pp. 191-194, 2008.

[22] C. Hendrix and W. Barfield, "The sense of presence within auditory virtual environments," Presence: Teleoperators & Virtual Environments, vol. 5, no. 3, pp. 290-301, 1996.