Asia Pacific Academy of Science Pte. Ltd.

Metaverse 2026, 7(3), 3676.
https://doi.org/10.54517/m3676

Article

The Effects of Spatial Sound on Immediate Memory Recall in a Web-based Virtual Reality Learning Environment

Vincent Russell[1], Billy O'Mahony[1], David Murphy[1,*]

[1] School of Computer Science & Information Technology, University College Cork, Cork, Ireland
***Corresponding author:** David Murphy, d.murphy@ucc.ie

CITATION

Russell V, O'Mahony B, Murphy D. The Effects of Spatial Sound on Immediate Memory Recall in a Web-based Virtual Reality Learning Environment. Metaverse. 2026; 7(3): 3676. https://doi.org/10.54517/m3676

ARTICLE INFO

Received: 1 May 2025
Accepted: 6 July 2026
Available online: 4 August 2026

Abstract: Virtual Reality Learning Environments (VRLEs) are increasingly adopted across educational domains due to their immersive and interactive capabilities. While prior research has largely focused on the visual and interactive affordances of Virtual Reality (VR), comparatively less is known about the influence of auditory modality on learning outcomes. In particular, the extent to which spatial audio affects information retention in immersive learning environments remains unclear. This study investigated whether spatial audio configuration influenced immediate memory recall performance within a WebXR-based VRLE, and whether this relationship was moderated by simulated room acoustic condition. Participants completed a short-story learning task within a controlled virtual classroom under two audio playback configurations and two room acoustic conditions. Immediate recall was assessed using recognition-based multiple-choice questions. A total of 39 participants completed the experiment, with 32 datasets retained following exclusion of invalid or uncontrolled submissions. Recall performance was analysed using a 2 × 2 mixed-design ANOVA, with spatial audio configuration as a within-subjects factor and room acoustic condition as a between-subjects factor. No statistically significant main effect of spatial audio configuration was observed, $F(1,30) = 0.053$, $p = .819$, $\eta p^2 = .002$, nor was there a significant main effect of room acoustics, $F(1,30) = 0.065$, $p = .801$, $\eta p^2 = .002$. No significant interaction was found, $F(1,30) = 1.607$, $p = .215$, $\eta p^2 = .051$. Sensitivity analysis indicated the design was powered to detect only relatively large effects ($f \approx 0.38$), suggesting smaller effects may have gone undetected. Further research with larger samples and more demanding tasks is required to clarify the educational value of spatial audio in immersive VR contexts.

Keywords: Virtual reality; Spatial sound; 3D audio; Web audio; Immediate memory recall; WebXR; Ambisonics; Learning environments

1. Introduction

Virtual Reality (VR) has emerged as a significant technology within educational research and practice, offering immersive and interactive environments that extend beyond the capabilities of conventional digital learning platforms. By enabling users to engage with simulated three-dimensional spaces through embodied interaction, VR can support forms of experiential learning that are difficult to replicate through traditional screen-based or classroom instruction. Educational applications of VR have been explored across disciplines including medicine, engineering, heritage studies, and the arts, where simulated environments allow learners to safely explore otherwise inaccessible, hazardous, or resource-intensive scenarios [1].

A defining characteristic of VR Learning Environments (VRLEs) is their inherently multimodal nature. Unlike conventional e-learning systems, VR integrates multiple sensory channels simultaneously, typically combining visual, auditory, and interactive spatial cues within a unified environment. This multimodal structure has been associated with increased learner engagement, improved spatial understanding, and stronger perceptions of presence and immersion [2,3]. These properties are frequently cited as potential mechanisms through which VR may enhance educational outcomes.

While much of the existing literature has focused on the visual and interactive affordances of VR, comparatively less attention has been given to the role of the auditory modality in VR learning contexts. In particular, spatial audio has been proposed as an important component of immersive system design, enabling sound sources to be perceived as originating from specific positions within three-dimensional space. By more closely approximating real-world auditory perception, spatial audio may increase environmental realism, improve attentional orientation, and support more effective encoding of information presented in virtual environments.

Within educational VR applications, however, the cognitive impact of spatial audio remains insufficiently understood. Although prior work has demonstrated that immersive VR environments can positively influence motivation, engagement, and spatial cognition [4-6], it is less clear whether increasing auditory realism through spatial sound confers measurable benefits for information retention or recall. This question is particularly relevant for WebXR-based educational systems, where implementation constraints often require developers to make trade-offs between technical complexity and pedagogical benefit.

Accordingly, the present study investigated whether spatial audio configuration influences immediate memory recall performance within a WebXR-based VRLE, and whether this effect varies as a function of simulated room acoustic condition.

Specifically, the study addressed the following research question: Does the use of spatial audio, relative to non-spatial audio playback, produce measurable differences in immediate recall performance within a VRLE, and is this relationship moderated by room acoustics?

To examine this, participants completed a recall-based learning task under differing audio playback configurations and simulated acoustic conditions, allowing the influence of auditory modality on immediate recall performance to be assessed experimentally.

1.1. Related Research

Spatial audio refers to sound reproduction techniques that enable auditory stimuli to be perceived as originating from specific locations within a three-dimensional environment. Common approaches include binaural rendering, ambisonics, and other forms of 3D audio processing that simulate directional and distance-based auditory cues [7]. Unlike conventional stereo or non-spatial playback, spatial audio introduces location-dependent information intended to more closely approximate real-world auditory perception.

Within immersive systems, spatial audio is frequently regarded as an important contributor to presence, environmental realism, and user orientation. By supporting directional awareness and improving auditory source localisation, spatial sound has the potential to influence attentional processes and cognitive performance within virtual environments.

A substantial body of prior research has examined the relationship between auditory stimuli and cognitive function, particularly with respect to short-term and immediate memory. Immediate memory refers to the temporary maintenance and manipulation of information required for tasks such as comprehension, reasoning, and recall [8]. Existing literature demonstrates that auditory input can meaningfully affect these processes, although the direction of this effect is highly context-dependent.

Numerous studies have shown that irrelevant or competing auditory stimuli can impair short-term recall performance, particularly in serial recall or speech-processing tasks. Both structured auditory interference (e.g., background speech) and unstructured environmental noise have been associated with reduced memory accuracy and increased cognitive load [9]. However, not all auditory stimulation is detrimental. Under certain conditions, low-level or non-verbal auditory inputs have been reported to produce neutral or facilitative effects on performance, depending on task complexity, individual cognitive differences, and familiarity with the auditory environment. For example, Söderlund *et al*. [10] reported that moderate white noise may improve task performance in individuals with attentional difficulties.

More specifically, recent research has explored the role of spatialised auditory cues in supporting cognitive performance. Spatial separation of auditory streams has been shown to improve speech intelligibility, reduce informational masking, and decrease memory interference in multi-source listening environments. Brungart and Simpson [11], for example, demonstrated that spatial separation of competing speakers improved speech comprehension and recall of target information. Similarly, Baldis [12] reported that spatialised audio improved task efficiency and memory performance in a desktop conferencing environment, likely by facilitating source separation and reducing cognitive demand.

Despite these reported benefits, the cognitive effects of spatial audio are not uniformly positive. Under more complex or information-dense auditory conditions, spatialisation may introduce additional sensory demands that compete for limited attentional resources. Leist *et al*. [13] reported that realistic binaural classroom noise impaired listening comprehension and increased perceived mental effort, suggesting that ecologically realistic spatial soundscapes may sometimes hinder rather than support cognitive performance. This finding is consistent with cognitive load theory, which proposes that increased sensory complexity can overload working memory resources when task demands are already substantial [14].

Within VR specifically, research investigating the relationship between spatial audio and cognitive performance remains comparatively limited. Existing VR studies have primarily focused on attention, presence, and perceptual performance rather than memory outcomes. Meghanathan *et al*. [15], for example, examined auditory cueing during visual search tasks in 360° virtual environments and found that binaural audio improved target detection speed and gaze efficiency relative to stereo playback. These findings suggest that spatial audio may enhance attentional orientation and information processing within immersive environments, potentially supporting downstream encoding and recall.

However, relatively few studies have directly examined whether spatial audio influences memory recall performance in immersive educational or learning-oriented VR environments. Furthermore, little is known about whether the effectiveness of spatial audio may depend on the acoustic characteristics of the virtual environment itself, such as the presence or absence of simulated room acoustics.

Accordingly, the present study addresses this gap by investigating whether positional spatial audio, with and without simulated room acoustic characteristics, influences immediate memory recall within a web-based VRLE. In doing so, this work extends existing research on auditory modality in immersive systems by examining not only the effect of spatial audio playback, but also its interaction with environmental acoustic realism in a controlled educational task.

2. Methods

2.1. Experimental Design

2.1.1. Platform Architecture

The randomised controlled trial (RCT) was conducted entirely within an online VRLE using open-source web technologies, following the framework proposed in detail by Russell *et al*. [16], which focused on the design, implementation, and deployment of the system. The present paper extends that work by reporting the empirical findings obtained using that platform and presents the corresponding statistical analysis, with particular emphasis on evaluating the effects of spatial audio and simulated room acoustics on immediate recall performance in a VRLE.

The platform was implemented using the WebXR Device API (XRDA)[1], enabling browser-based deployment across desktop computers, mobile devices, and VR headsets without requiring third-party plugins. This cross-platform architecture supported a “build once, deploy everywhere” approach suitable for scalable educational and experimental delivery [17-20].

The VRLE was developed using the Three.js[2] JavaScript library, which provided the rendering framework for the browser-based 3D environment [21]. Spatial audio functionality was implemented using the Resonance Audio SDK[3], built upon the Web Audio API[4]. Resonance Audio was selected due to its support for head-related transfer functions (HRTFs), binaural rendering, and room-based acoustic simulation, which were central to the experimental manipulation.

As the experiment was delivered entirely through the head-mounted display’s (HMD) web browser, asset size and computational efficiency were prioritised throughout development. Performance was monitored continuously to maintain rapid loading times and a minimum frame rate of 60 FPS, thereby reducing the likelihood of motion sickness symptoms such as nausea, discomfort, and disorientation.

The experimental platform was developed as a WebXR application to facilitate both laboratory-based and remote deployment. At the time of writing, the platform is no longer publicly accessible and is therefore unavailable for independent replication or reuse in its original deployed form.

2.1.2. Virtual Environment Optimisation

A virtual classroom was selected as the experimental setting. The classroom environment, including furniture and presentation objects (e.g., podium, projector, speakers), was modelled in Blender and exported as a single GLB asset for efficient loading within Three.js.

To reduce computational overhead, mesh optimisation procedures were applied, including removal of non-visible polygon faces and polygon reduction using Blender’s decimate modifier. A unified UV map and shared texture map were generated to minimise texture loading requirements.

[1] https://www.w3.org/TR/webxr/
[2] https://threejs.org
[3] https://resonance-audio.github.io/resonance-audio/
[4] https://www.w3.org/TR/webaudio-1.1/

Additional optimisation was achieved through object instancing. For example, a single desk-chair model was loaded once and duplicated throughout the classroom using object cloning, thereby reducing memory demands.

Lighting calculations were minimised to preserve rendering performance. Scene illumination consisted of a single ambient light source without shadow casting, while additional lighting and shading information was precomputed in Blender and baked into texture maps. For the classroom floor, a separate lightmap containing baked illumination and shading was combined with a tiled seamless wood texture, reducing texture memory requirements while preserving visual realism. The completed virtual classroom is shown in **Figure 1**.

Figure 1. The completed virtual classroom of the VRLE.

2.1.3. Spatial Audio Configuration

All audio stimuli and experimental audio conditions were processed using the Resonance Audio SDK to ensure consistent rendering across conditions. The experiment exclusively used verbal speech recordings captured in an acoustically controlled recording studio to eliminate unwanted reverberation and ensure compatibility with the simulated acoustics of the VRLE.

Dialogue stimuli were recorded in mono WAV format (48 kHz, 16-bit) within a recording studio to remove any undesired reverberation and extraneous background noise. These dry recordings were subsequently processed in Resonance Audio to generate acoustically congruent wet audio through simulated reflections and reverberation. Within Resonance Audio, a *room model* was configured measuring 11 m × 8.5 m × 3 m. Acoustic material coefficients were selected to correspond to the classroom's visual materials, including acoustic ceiling tiles, glass, painted concrete walls, and wooden flooring.

In accordance with established VR audio design practices [22], speech stimuli were spatially positioned at the head location of an animated (limited to hand gestures) non-player character (NPC) presenter (**Figure 2**), located at a fixed position in the front-right area of the classroom. Audio spatialisation was applied via generalised HRTF convolution according to source coordinates, with activation contingent on the participant's randomised experimental condition.

Figure 2. The fixed position NPC from which presentation audio emanated.

2.1.4. Immediate Memory Assessment

Immediate memory performance was assessed using an adapted version of the Logical Memory (LM) test [23], a widely used subtest of the Wechsler Memory Scales [24] and a reliable measure of narrative episodic memory in individuals aged 16–90 years [25]. The LM test consists of two short stories, *Anna Thompson* and *Joe Garcia*, comprising 65 and 85 words respectively, each containing 25 scorable information units ("bits"). Under standard clinical administration, an examiner orally presents each story once at a controlled pace, after which participants are asked to freely recall as many details as possible. Scoring is based on the number of correctly recalled information units. Following a delay of approximately 25 minutes, delayed recall is assessed.

As delayed recall was impractical within a self-administered WebXR environment, only immediate memory performance was evaluated. The examiner was replaced by the NPC, which delivered consistent audio playback of the stories. Free recall was replaced with in-VR multiple-choice questions (MCQs). Prior research has successfully supplemented LM with multiple-choice recognition items to differentiate storage from retrieval processes [26,27]. In adapting LM for immersive VR, a recognition-based MCQ format employing competitive distractors [28] was selected to reduce construct-irrelevant variance (CIV) associated with speech production, manual scoring, and VR interface-related demands. This design also enabled automated scoring while preserving participant presence and flow by minimising breaks in presence (BIP). The approach is also consistent with prior work demonstrating the validity of VR for episodic memory assessment [29,30].

2.1.5. In-VR Questionnaire Delivery

In-VR questionnaires (inVRQs) were employed due to the autonomous nature of the web-based experiment. Relative to traditional paper-based methods, inVRQs offer advantages including improved data security, reduced administrative burden, and lower risk of transcription errors during data collation. They also minimise

BIP [31], are less intrusive, and can yield more reliable participant responses [32]. From the participant perspective, inVRQs reduce disruptions to immersion and decrease the likelihood of adverse reactions that may compromise response validity [33]. Questionnaires were displayed on semi-transparent 2D planar surfaces using the CanvasUI[5] library (**Figure 3**). This interface approach has previously been shown to be faster and less frustrating than 3D object-based interaction methods for in-VR questionnaire completion [34].

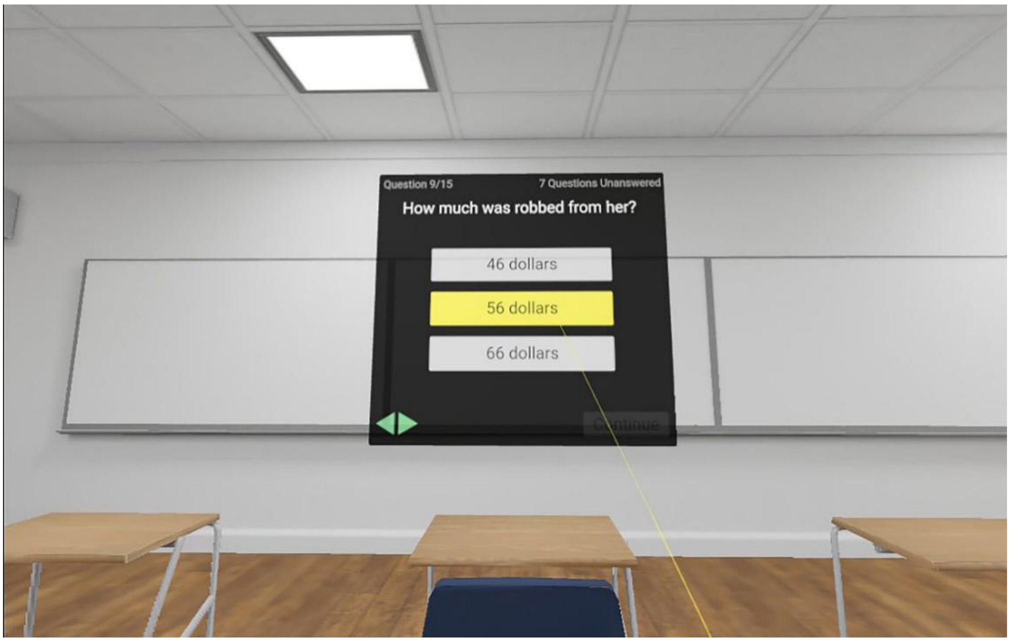

Figure 3. An example of an inVRQ used to capture test answers.

2.2. Participants

Participants were recruited for both laboratory-based and remote deployment of the experimental system. However, only data collected under controlled laboratory conditions were retained for final analysis to ensure experimental consistency and data integrity. Consequently, all analysed experimental sessions were supervised and conducted within the VR laboratory at University College Cork, providing a standardised testing environment across participants. Eligibility criteria required participants to be aged 18 years or older. No restrictions were placed on gender, technical background, or prior experience with VR technologies.

To ensure consistency of hardware and audiovisual presentation, all participants completed the experiment using an identical system configuration consisting of a Meta Quest 2 HMD with integrated headphones. This configuration enabled standardised delivery of the HRTF-based binaural spatial audio implementation within the VRLE.

Audio playback volume was fixed at 90% system output for all participants to ensure consistent exposure to the spatialised audio stimuli across experimental conditions.

2.3. Process

Following placement of the HMD, participants accessed the experimental website via the device's integrated web browser, thereby initiating the experimental procedure illustrated in **Figure 4**.

[5] https://github.com/NikLever/CanvasUI

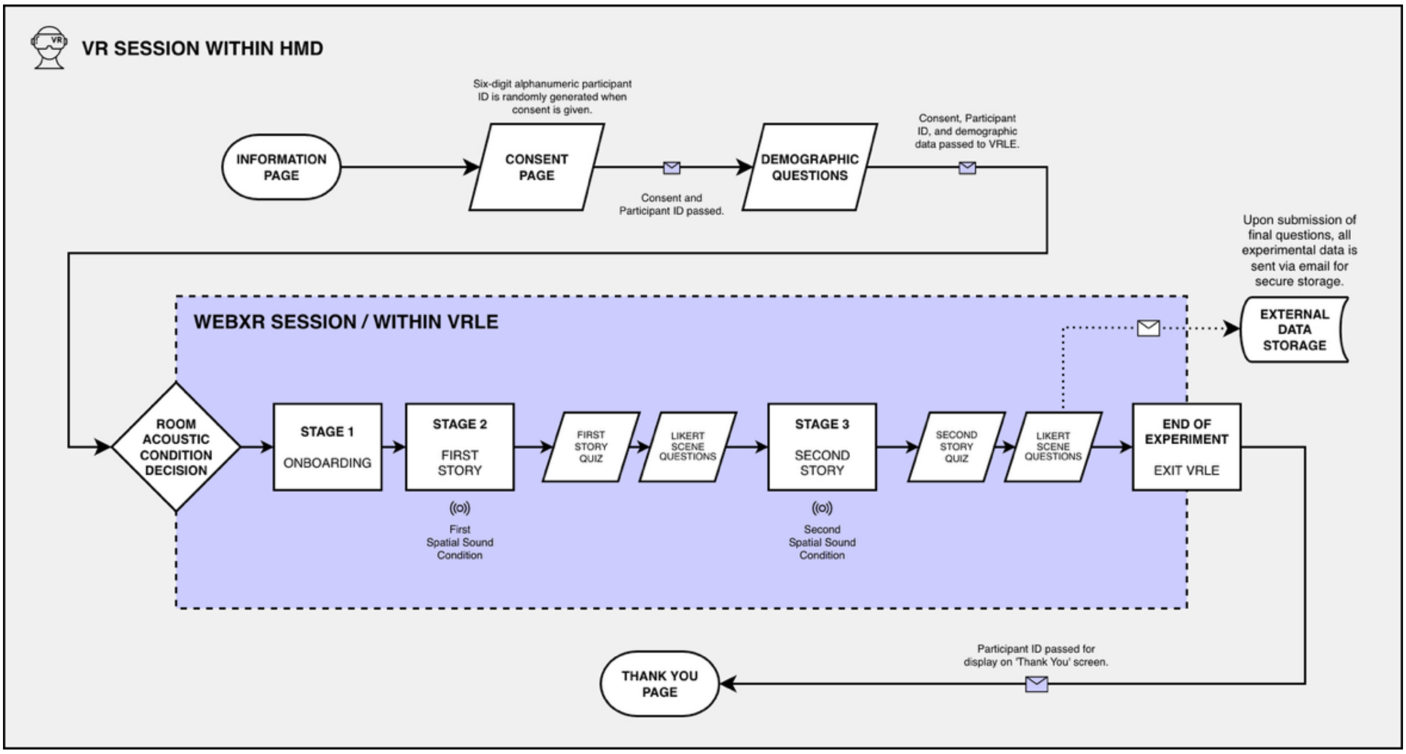

Figure 4. The experimental procedure flow diagram.

Participants were first presented with an information page outlining the study objectives and procedures. Those willing to participate proceeded to the digital consent form. Upon providing consent, each participant was assigned a randomly generated six-character alphanumeric Participant ID. This identifier was displayed on screen and known only to the participant, enabling later withdrawal of data if requested. The Participant ID was automatically linked to all submitted responses and displayed again upon completion of the experiment. Participant data were anonymised from the outset, and no personally identifiable information was collected at any stage.

Following consent, participants completed a brief demographic questionnaire presented as a simplified web form with large, easily selectable interface elements optimised for VR interaction. Upon submission, the virtual classroom environment was loaded, and participants were reminded to remain seated and ensure audio output was enabled on the HMD.

At this stage, the system performed uniform random assignment to determine: (1) the participant's room acoustics condition, (2) the order of spatial audio presentation configuration, and (3) the sequence of LM test narratives. Once assigned, the room acoustics condition remained fixed throughout the entire experiment.

Participants were allocated to one of two acoustic groups. Group 0 experienced the virtual classroom with no acoustic treatment (dry sound), with room acoustics disabled. Group 1 experienced the same environment with room acoustics enabled, simulating a natural reverberant classroom space. Within their assigned acoustic group, all participants completed both LM test stories under two spatial audio presentation configurations:

1. Non-spatial audio, in which sound remained head-locked (stereo audio fixed relative to head orientation); and

2. Spatial audio, in which sound was anchored to a fixed position within the virtual classroom and dynamically updated according to participant head tracking.

Table 1 summarises the terminology used for group allocation and audio presentation configurations. The order of story presentation and the sequence of audio conditions were randomised for each participant.

Table 1. Grouping and audio presentation configurations.

Group 0	Room Acoustics OFF
Group 1	Room Acoustics ON
Non-spatial	Spatialisation OFF / Head-locked Audio
Spatial	Spatialisation ON / Head-tracked Audio

Table 2. The fundamental experiment matrix without any randomisation.

	Non-spatial	**Spatial**
Group 0	Story 1 Test Score (15 Qs)	Story 2 Test Score (15 Qs)
Group 1	Story 1 Test Score (15 Qs)	Story 2 Test Score (15 Qs)

Table 2 presents the fundamental experimental matrix prior to randomisation. Through randomisation of room acoustics, audio presentation configuration order, and narrative sequence, participants experienced one of eight possible experimental variations, detailed in **Table 3**.

Table 3. The eight possible experiment permutations.

	Room Acoustics	**Story 1**	**Story 2**
1	**Off**	Joe – Non-spatial	Anna – Spatial
2	**Off**	Anna – Non-spatial	Joe – Spatial
3	**Off**	Joe – Spatial	Anna – Non-spatial
4	**Off**	Anna – Spatial	Joe – Non-spatial
5	**On**	Joe – Non-spatial	Anna – Spatial
6	**On**	Anna – Non-spatial	Joe - Spatial
7	**On**	Joe – Spatial	Anna – Non-spatial
8	**On**	Anna – Spatial	Joe – Non-spatial

Upon entering the immersive virtual classroom, participants were positioned in a seated location at a desk placed centrally within the environment. A welcome audio message introduced the VRLE and encouraged participants to spend time familiarising themselves with the virtual space before proceeding. Two instructional information panels were presented to explain interaction controls and selection methods used throughout the experiment. This onboarding phase was self-paced and intentionally untimed, allowing participants to acclimatise to the environment and progress only when comfortable.

Following onboarding, participants proceeded to the primary experimental task involving the adapted LM assessments. The NPC presenter, positioned at the front-right of the classroom, delivered a brief explanation of the testing procedure.

Participants then listened to the first allocated LM story under the first assigned audio presentation configuration. Immediately following story presentation, participants completed fifteen MCQs assessing recall of narrative scoring "bits".

After the memory assessment, participants completed nine post-exposure subjective rating items using 5-point Likert scales ranging from *Strongly Disagree* to *Strongly Agree* (**Figure 5**). Questionnaire items were adapted from the Igroup Presence Questionnaire (IPQ) [35] and Spatial Audio Quality Inventory (SAQI) [36], and assessed perceived presence, audio characteristics, and auditory realism associated with the preceding narrative presentation. Following completion of these items, participants listened to the second LM narrative presented under the alternate audio presentation configuration. This was again followed by fifteen MCQs and the same nine Likert-scale questionnaire items.

One questionnaire item presented after each story was excluded from subsequent analysis. This item assessed awareness of external real-world events; however, due to uncontrolled and variable ambient sound conditions outside the laboratory setting, responses were considered insufficiently standardised across participants for valid comparison. The remaining eight-item questionnaire demonstrated good internal consistency reliability, with a Cronbach's alpha of $\alpha = .810$. Given that the questionnaire met the conventional threshold for high internal consistency ($\alpha > .80$), responses were deemed sufficiently reliable for further parametric and non-parametric comparative analyses.

Upon completion of all experimental tasks, participants were thanked for their contribution, and their Participant ID was displayed again to facilitate potential future withdrawal of data.

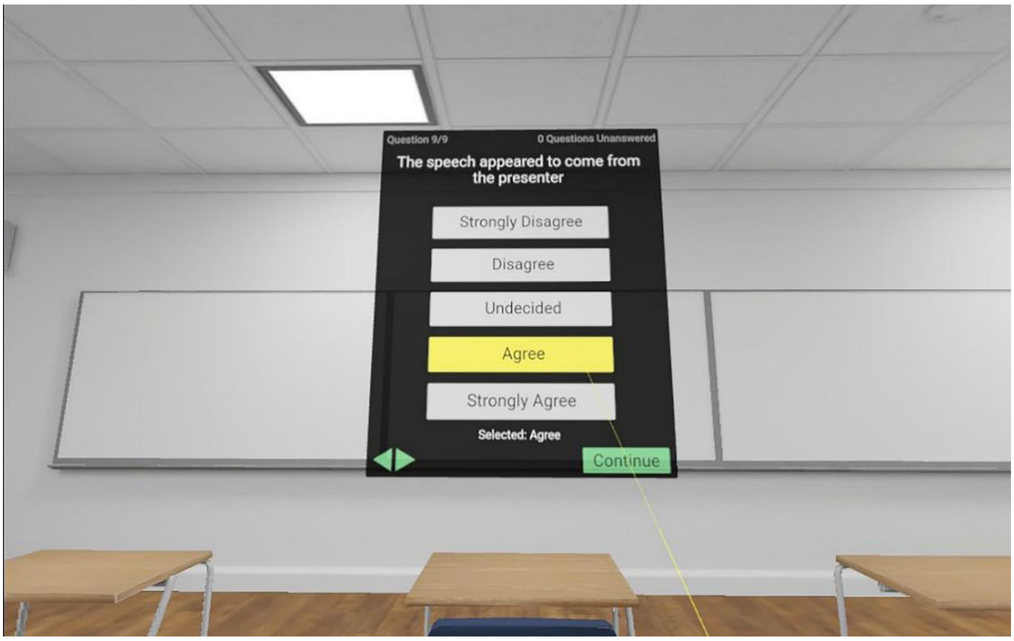

Figure 5. An example of a 5-point Likert scale item to record preferences.

The system automatically collated all submitted data, including demographic responses, MCQ responses, and post-story questionnaire data. These data were transmitted via secure email to the researcher upon completion of the experiment for encrypted storage and subsequent analysis.

2.4. Statistical Analysis

Statistical analyses were conducted using IBM SPSS Statistics (version

29.0.2.0 (20)). Prior to analysis, participant data were screened for validity and completeness. Data from participants affected by hardware-related recording issues were excluded, as were responses collected outside controlled laboratory conditions to ensure experimental consistency and data integrity.

Immediate recall performance was analysed using a 2 × 2 mixed-design analysis of variance (ANOVA). Spatial audio condition was treated as a within-subjects factor with two levels corresponding to the two playback conditions, while the room acoustic condition was treated as a between-subjects factor with two levels representing the presence or absence of simulated room acoustics. The mixed ANOVA was used to evaluate:

1. the main effect of spatial audio condition on recall performance,
2. the main effect of room acoustic condition, and
3. the interaction between spatial audio condition and room acoustic condition.

Prior to inferential analysis, assumptions for mixed ANOVA were assessed. Normality of the dependent variable was evaluated using Shapiro–Wilk tests in conjunction with visual inspection of Q–Q plots. Homogeneity of variances across groups was assessed using Levene's test for each combination of factor levels. Potential outliers were examined through visual inspection of the data distributions and by calculating standardised scores, with all retained observations falling within an acceptable threshold of ±3.

Descriptive statistics, Estimated Marginal Means, and 95% confidence intervals were computed for all experimental conditions. Where appropriate, Bonferroni-adjusted pairwise comparisons were conducted to assess mean differences between factor levels while controlling for multiple comparisons.

Statistical significance was evaluated using an alpha level of $\alpha = .05$. Effect sizes are reported as partial eta squared (ηp^2) to quantify the magnitude of observed effects.

As the within-subjects factor contained only two levels, no correction for violations of sphericity was required.

To assess the sensitivity of the experimental design, a post hoc sensitivity analysis was conducted in G*Power[6] using an alpha level of .05, desired statistical power of .80, and the final analysed sample size. This analysis was used to estimate the minimum detectable effect size for the interaction term under the present mixed factorial design.

Regarding the post-story questionnaires, as individual item differences showed significant departures from normality according to Shapiro-Wilk testing, and paired responses were ordinal in nature, the non-parametric Wilcoxon signed-rank test was regarded as methodologically appropriate for subsequent inferential analyses.

3. Results

3.1. Participant Characteristics and Data Screening

A total of 39 participants completed the experimental procedure. Prior to analysis, three participants were excluded due to hardware-related issues that compromised data capture accuracy. An additional four participants were excluded because they completed the study remotely outside controlled laboratory conditions, preventing consistency of environmental and equipment variables. Following these exclusions, data from 32 participants (N = 32) who completed the study under

[6] https://www.psychologie.hhu.de/arbeitsgruppen/allgemeine-psychologie-und-arbeitspsychologie/gpower

supervised laboratory conditions were retained for final analysis.

Of the final sample, 46.9% (*n* = 15) identified as male, 50.0% (*n* = 16) as female, and 3.1% (*n* = 1) as other. Participants were predominantly younger adults, with 81.3% (*n* = 26) aged under 30 years. The sample was largely student-based, with 84.4% (*n* = 27) identifying as university students, and all participants (100%, *n* = 32) reporting education to university level.

Given the VR-based nature of the experiment and its intended relevance to immersive educational applications, prior familiarity with VR technology was also assessed. Demographic responses indicated that 62.5% (*n* = 20) of participants reported little to no previous VR experience, while the remaining participants reported some prior familiarity with VR systems.

Participants were randomly assigned to one of two room acoustic conditions, resulting in an equal allocation of 16 participants per group. The breakdown of the sample population is outlined in **Table 4**.

Table 4. Sample population breakdown

Population Total	N = 32
Group 0	N0 = 16
Group 1	N1 = 16

3.2. Recall Performance

Recall performance was analysed using a 2 × 2 mixed-design ANOVA, with spatial audio condition entered as a within-subjects factor and room acoustic condition entered as a between-subjects factor.

Descriptive statistics indicated similar recall performance across spatial conditions. Mean recall scores were M = 10.84, SE = 0.41 for non-spatial audio and M = 10.97, SE = 0.32 for spatial audio.

The analysis revealed no significant main effect of spatial audio condition, $F(1, 30) = 0.053$, $p = .819$, $\eta p^2 = .002$, indicating no statistically detectable overall difference in recall performance between spatial audio conditions.

There was likewise no significant main effect of room acoustic condition, $F(1, 30) = 0.065$, $p = .801$, $\eta p^2 = .002$, indicating that overall recall performance did not differ significantly between acoustic environments.

The interaction between spatial audio and room acoustic conditions was also not statistically significant, $F(1, 30) = 1.607$, $p = .215$, $\eta p^2 = .051$, suggesting that the effect of spatial audio configuration on recall did not significantly vary as a function of room acoustics.

Bonferroni-adjusted pairwise comparisons further supported these findings, showing no significant difference between spatial audio conditions (mean difference = -0.125, SE = 0.542, 95% CI [-1.233, 0.983], $p = .819$) and no significant difference between room acoustic conditions (mean difference = 0.125, SE = 0.492, 95% CI [-0.880, 1.130], $p = .801$).

Estimated Marginal Means for each condition are visualised in **Figure 6**. Although a modest crossover pattern is visible between room acoustic conditions across spatial audio conditions, confidence intervals overlap substantially and the interaction effect was not statistically significant.

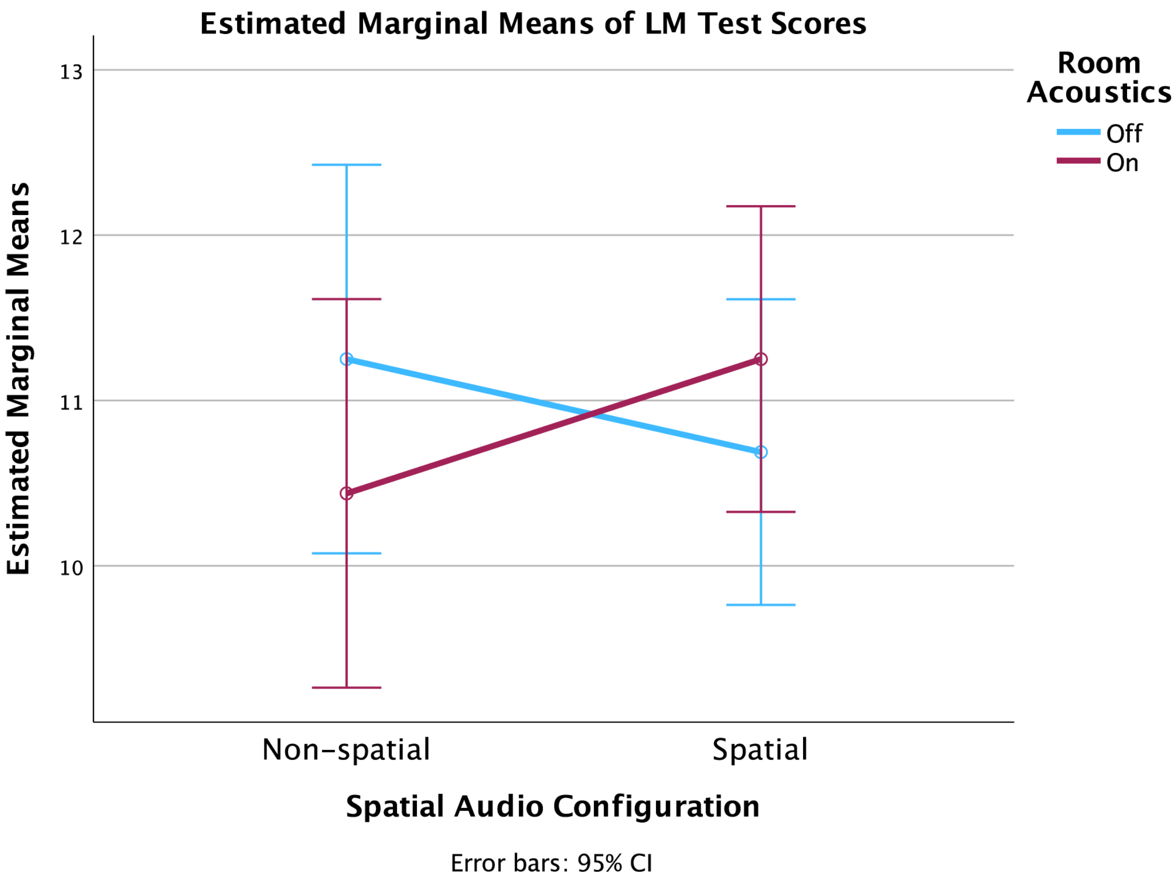

Figure 6. Estimated Marginal Means for immediate recall performance across spatial audio conditions under each room acoustic condition. Error bars represent 95% confidence intervals.

3.3. Statistical Sensitivity

To assess the statistical sensitivity of the present design, a sensitivity analysis was conducted in G*Power using $\alpha = .05$, desired power = .80, total sample size = 32, and the observed correlation among repeated measures.

The analysis indicated that the present design was sufficiently powered to detect interaction effects of approximately $f = 0.38$ or larger. According to conventional benchmarks, this corresponds approximately to effects approaching the large range, indicating that the study was primarily sensitive to relatively large effects only.

Accordingly, while no statistically significant effects of spatial audio condition, room acoustic condition, or their interaction were detected, smaller effects may not have been detectable under the present sample size and experimental design.

3.4. Subjective Questionnaire Measures

A series of Wilcoxon signed-rank tests were conducted for the individual eight-item post-story questionnaires to evaluate distributional shifts between the non-spatial and spatial audio conditions under each room acoustic condition. The Likert-scale questionnaire items assessed perceived presence, awareness of real-world surroundings, sound quality, immersion, environmental consistency, realism, speech clarity, and source localisation.

For the room acoustics enabled condition, no statistically significant differences were observed between spatial and non-spatial audio on any questionnaire measure (all $p > .05$). Similarly, under the room acoustics disabled condition, no significant differences were observed for perceived presence, awareness of real-world surroundings, sound quality, immersion, environmental consistency, realism, or speech clarity (all $p > .05$).

The only questionnaire item that achieved marginal statistical significance was the statement “the speech appeared to come from the presenter” under the room acoustics disabled condition ($Z = -1.980$, $p = .048$), with a directional shift in means (M_Spatial = 2.19 vs. M_Non-spatial = 2.88) indicating that participants perceived stronger source localisation when spatial audio was disabled. The result should be interpreted with caution due to its proximity to the alpha boundary ($\alpha = .05$) and small overall sample size. This effect had no corresponding observation when room acoustics were enabled ($Z = -0.877$, $p = .380$) and did not pass the Bonferroni correction for multiple comparisons, leading to its classification as statistically unreliable.

4. Discussion

The present study investigated the effect of spatial audio configuration on immediate recall performance in a WebXR-based VRLE and the interaction between these factors with simulated room acoustic conditions. No statistically significant main effect of spatial audio condition was observed on immediate recall performance across the present sample. Similarly, there was no statistically significant interaction between spatial audio and room acoustic conditions, and the room acoustic condition did not significantly affect recall.

The findings indicate that the immediate recall performance did not improve with the use of spatial audio under the experimental conditions employed. These findings, however, should be interpreted cautiously. Sensitivity analysis indicated that the current design was sufficiently powered to detect only relatively large interaction effects ($f \approx 0.38$ or greater), suggesting that smaller effects may not have been detected with the available sample size.

This limitation is particularly important when interpreting the non-significant interaction between spatial audio and room acoustic conditions. This interaction represents arguably the most theoretically important effect examined in the present study, as it directly addresses whether the influence of spatial audio depends upon the acoustic characteristics of the virtual environment. Although this interaction was not statistically significant, the observed effect size ($\eta p^2 = .051$; $f \approx 0.23$) fell below the study's estimated detection threshold ($f \approx 0.38$). Consequently, the present sample size and experimental design may have lacked sufficient sensitivity to detect interaction effects of this magnitude and should not be interpreted as ruling out smaller interaction effects that may emerge in larger samples.

Lack of significant effects could also be attributable to characteristics of the learning task itself. The study was initially designed to be conceptually similar to immediate verbal recall paradigms, but the final assessment format was a multiple-choice recognition task instead of free recall. Tasks that require recognition are generally less demanding on retrieval than recall tasks because the target information is cued externally rather than reconstructed internally. Thus, the cognitive load of the task may not have been high enough to allow for subtle differences in modality to be detected.

While no participant achieved a perfect score, the overall performance was relatively high, with four participants scoring 14/15 and an overall mean score of approximately 11. This indicates that the task may have been less sensitive to finer-grained differences in encoding or retrieval performance. As such, the assessment format may have limited the ability to detect subtle cognitive effects due to auditory condition.

A further limitation concerns the implementation of spatial audio itself. In the present study, Resonance Audio was employed, which utilises HRTFs from the SADIE[7] binaural filter library to simulate spatial hearing. These HRTFs are of high quality, but they are not personalised and thus are an approximation of the average human auditory perception. It is well known that generic HRTFs can cause perceptual errors such as front-back confusions, elevation errors, decreased externalisation and weakened localisation precision [37-39]. These artefacts can decrease the perceptual salience of spatial audio cues and therefore the potential downstream cognitive benefits.

The subjective questionnaire data did not provide robust evidence that participants perceived meaningful differences between the spatial and non-spatial audio conditions. Although one source-localisation item reached nominal statistical significance, with participants reporting stronger agreement that speech appeared to originate from the presenter under the non-spatial condition when room acoustics were disabled, this effect did not survive correction for multiple comparisons and should therefore be interpreted cautiously. Nevertheless, the finding is not entirely inconsistent with established phenomena in spatial acoustics and audiovisual perception. Unlike spatialised audio, standard non-spatial stereo playback over headphones delivers a dry, highly proximate sound signal directly to the listener, which can induce in-head localisation (IHL) [40]. Additionally, by having no competing environmental or spatial cues, non-spatialised audio can strengthen the perceived association between sound and a visually prominent target. In the present study, the NPC presenter may have acted as a dominant visual anchor, producing a ventriloquist effect [41] whereby participants attributed the speech signal to the presenter despite the absence of explicit spatial cues. Conversely, the spatial audio condition relied on generic HRTFs, which are known to introduce “localisation blur” [42], front-back confusions, reduced externalisation, and spatial diffusion [43,44]. As a result, the isolated questionnaire finding may reflect limitations of non-personalised binaural rendering rather than a genuine advantage of non-spatial audio, although further investigation would be required to determine whether this effect is reliable.

One possible reason for the absence of significant effects from the questionnaire data is that these subjective ratings were taken immediately after each story presentation, requiring participants to evaluate conditions in isolation rather than comparatively. This could have caused ceiling or floor effects or reduced sensitivity to perceived differences between spatial audio playback conditions. Future studies may benefit from collecting preference or perceptual comparison measures after both conditions have been experienced, allowing for more direct relative judgements.

The effectiveness of the auditory manipulation could also have been affected by the hardware configuration. The Meta Quest 2's off-ear speakers are generally suitable for binaural playback by allowing natural interaction between sound waves and the listener’s pinnae, it is uncertain how effective they are compared to closed-back or over-ear isolating headphones. It is therefore possible that the perceptual distinctiveness of the spatial audio conditions may have been restricted by the hardware configuration.

[7] https://www.york.ac.uk/sadie-project/GoogleVRSADIE.html

Furthermore, the experimental setting might not have been conducive to maximising the perceptual advantages of spatial audio. For instance, the presence of a visually and aurally dynamic sound source (e.g., a moving NPC or animated speaker) within the virtual classroom could have improved the perceptibility of source localisation and increased the sense of spatial immersion. Spatialisation may have been less effective in the present study, due to the relatively static presentation.

Overall, these findings indicate that spatial audio may not meaningfully influence immediate recognition-based recall performance in relatively simple WebXR learning tasks under the present conditions. This should not be interpreted as evidence that the pedagogical relevance of auditory realism is not relevant to VRLEs in general. Spatial audio may have more pronounced effects in situations with higher attentional load, more complex auditory scenes, delayed retention measures, spatial learning tasks, or when source localisation is directly relevant to task performance.

Future studies should therefore seek to replicate the present study with larger samples, more statistically sensitive designs, and more challenging cognitive tasks. Comparisons of personalised and non-personalised HRTFs, alternative headphone configurations, delayed recall paradigms, and more ecologically rich spatial learning scenarios may help to better understand when and how auditory realism meaningfully influences learning outcomes in immersive virtual environments.

A significant advantage of learning environments based on WebXR is their capability for extensive remote deployment utilising standard consumer hardware. While the current study initially supported both laboratory and remote participation, only the data collected under controlled laboratory conditions were retained for analysis to ensure consistency in the experiment and the integrity of the data. As a result, the benefits of scalability offered by WebXR were not fully explored in this study. Future research could more effectively utilise remote WebXR deployment while maintaining data quality by implementing additional standardisation and validation protocols. These may include logging participant hardware specifications (e.g., headset model, audio output device, browser version, and operating system), enforcing predefined audio volume levels, confirming headphone usage, recording environmental metadata, and integrating automated data quality assessments. Such strategies would mitigate variability caused by heterogeneous participant equipment and uncontrolled listening environments, especially in studies focused on spatial audio perception. The capacity to deliver immersive learning experiences remotely while preserving methodological rigour is a crucial direction for future WebXR research and may facilitate significantly larger and more diverse participant samples than what is typically possible through laboratory-based studies alone.

In summary, no statistically significant evidence was found to suggest that spatial audio configuration or room acoustic condition affected immediate recall performance in the present WebXR-based VRLE. However, given the study's limited sensitivity and the methodological limitations mentioned above, further research is required before firm conclusions can be drawn regarding the educational value of spatial audio in immersive learning environment contexts.

Author contributions: Conceptualization, VR and DM; methodology, VR and DM; software, VR and BOM; validation, VR, BOM and DM; formal analysis, VR; investigation, VR; resources, VR; data curation, VR; writing—original draft preparation, VR; writing—review and editing, VR and DM; visualization, VR; supervision, DM; project administration, VR. All authors have read and agreed to the published version of the manuscript.

Ethics: This study received institutional ethics approval. Social Research and Ethics Committee (SREC) 2022-003.

Conflict of interest: The authors declare no conflict of interest.

Funding: This research received no external funding.

References

1. Radianti J, Majchrzak TA, Fromm J, et al. A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda. Computers & Education. 2020,147:103778. http://doi.org/10.1016/j.compedu.2019.103778
2. Freina L, Ott M. A literature review on immersive virtual reality in education: State of the art and perspectives. In: The International Scientific Conference eLearning and Software for Education. 2015. Available online: https://www.academia.edu/download/93940955/eLSE_202015_20Freina_20Ott_20Paper.pdf (accessed on 18 April 2025)
3. Makransky G, Lilleholt L. A structural equation modeling investigation of the emotional value of immersive virtual reality in education. Educational Technology Research and Development. 2018,66(5):1141–1164. http://doi.org/10.1007/s11423-018-9581-2
4. Gittinger M, Wiesche D. Systematic review of spatial abilities and virtual reality: The role of interaction. Journal of Engineering Education. 2024,113(4):919–938. http://doi.org/10.1002/jee.20568
5. Cheng KH, Lee SWY, Hsu YT. The roles of epistemic curiosity and situational interest in students' attitudinal learning in immersive virtual reality environments. Journal of Educational Computing Research. 2023,61(2):494–519. http://doi.org/10.1177/07356331221121284
6. Lin XP, Li BB, Yao ZN, et al. The impact of virtual reality on student engagement in the classroom–a critical review of the literature. Frontiers in Psychology. 2024,15. http://doi.org/10.3389/fpsyg.2024.1360574
7. Begault DR, Trejo LJ. 3-D sound for virtual reality and multimedia. 2000. Available online: https://ntrs.nasa.gov/citations/20010044352 (accessed on 26 April 2025)
8. Baddeley A. The episodic buffer: A new component of working memory? Trends in Cognitive Sciences. 2000,4(11):417–423. http://doi.org/10.1016/S1364-6613(00)01538-2
9. Kämpfe J, Sedlmeier P, Renkewitz F. The impact of background music on adult listeners: A meta-analysis. Psychology of Music. 2011,39(4):424–448. http://doi.org/10.1177/0305735610376261
10. Söderlund GB, Sikström S, Loftesnes JM, et al. The effects of background white noise on memory performance in inattentive school children. Behavioral and Brain Functions. 2010,6(1):55. http://doi.org/10.1186/1744-9081-6-55
11. Brungart DS, Simpson BD. The effects of spatial separation in distance on the informational and energetic masking of a nearby speech signal. The Journal of the Acoustical Society of America. 2002,112(2):664–676. http://doi.org/10.1121/1.1490592
12. Baldis JJ. Effects of spatial audio on memory, comprehension, and preference during desktop conferences. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems. 2001:166–173. http://doi.org/10.1145/365024.365092
13. Leist L, Lacroix A, Grund M, et al. Differential effects of task-irrelevant monaural and binaural classroom scenarios on children's and adults' speech perception, listening comprehension, and visual–verbal short-term memory. International Journal of Environmental Research and Public Health. 2022,19(23):15998. http://doi.org/10.3390/ijerph192315998
14. Lavie N. Distracted and confused?: Selective attention under load. Trends in Cognitive Sciences. 2005,9(2):75–82. http://doi.org/10.1016/j.tics.2004.12.004
15. Meghanathan RN, Ruediger-Flore P, Hekele F, et al. Spatial sound in a 3D virtual environment: All bark and no bite? Big Data and Cognitive Computing. 2021,5(4):79. http://doi.org/10.3390/bdcc5040079
16. Russell V, Murphy D, Neff F. The design of an experiment to evaluate the effect of spatial sound on memory recall in a virtual reality learning environment. In: Proceedings of the 33rd European Conference on Cognitive Ergonomics. 2022:1–5. http://doi.org/10.1145/3552327.3552346
17. Maclntyre B, Smith TF. Thoughts on the future of WebXR and the immersive web. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE; 2018:338–342. http://doi.org/10.1109/ISMAR-Adjunct.2018.00099

18. Engberg M, Bolter JD, Maclntyre B. RealityMedia: An experimental digital book in WebXR. In: 2018 IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct). IEEE; 2018:324–327. http://doi.org/10.1109/ISMAR-Adjunct.2018.00096
19. Fanini B, Ferdani D, Demetrescu E. Temporal lensing: An interactive and scalable technique for Web3D/WebXR applications in cultural heritage. Heritage. 2021,4(2):710–724. http://doi.org/10.3390/heritage4020040
20. Rodríguez FC, Dal Peraro M, Abriata LA. Democratizing interactive, immersive experiences for science education with WebXR. Nature Computational Science. 2021,1(10):631–632. http://doi.org/10.1038/s43588-021-00142-8
21. Dirksen J. Learning Three.js: The JavaScript 3D Library for WebGL. Packt Publishing Ltd; 2013.
22. Zhang W, Samarasinghe PN, Chen H, et al. Surround by sound: A review of spatial audio recording and reproduction. Applied Sciences. 2017,7(5):532. http://doi.org/10.3390/app7050532
23. Rabin LA, Paolillo E, Barr WB. Stability in test-usage practices of clinical neuropsychologists in the United States and Canada over a 10-year period: A follow-up survey of INS and NAN members. Archives of Clinical Neuropsychology. 2016,31(3):206–230. http://doi.org/10.1093/arclin/acw007
24. Wechsler D. WMS-IV: Wechsler Memory Scale. Pearson; 2009.
25. Sullivan K. Alternate forms of prose passages for the assessment of auditory–verbal memory. Archives of Clinical Neuropsychology. 2005,20(6):745–753. http://doi.org/10.1016/j.acn.2005.04.006
26. Gass CS. A procedure for assessing storage and retrieval on the Wechsler Memory Scale — Revised. Archives of Clinical Neuropsychology. 1995,10(5):475–487. http://doi.org/10.1016/0887-6177(95)98192-G
27. Fastenau PS. An elaborated administration of the Wechsler Memory Scale-Revised. The Clinical Neuropsychologist. 1996,10(4):425–434. http://doi.org/10.1080/13854049608406703
28. Little JL, Frickey EA, Fung AK. The role of retrieval in answering multiple-choice questions. Journal of Experimental Psychology: Learning, Memory, and Cognition. 2019,45(8):1473–1485. http://doi.org/10.1037/xlm0000638
29. Sauzéon H, et al. The use of virtual reality for episodic memory assessment. Experimental Psychology. 2012,59(2):99-108. http://doi.org/10.1027/1618-3169/a000131
30. Höller Y, Höhn C, Schwimmbeck F, et al. A virtual reality paradigm to assess episodic memory: Validation-dataset for six parallel versions and a structured behavioral assessment. Data in Brief. 2020,29:105279. http://doi.org/10.1016/j.dib.2020.105279
31. Slater M, Brogni A, Steed A. Physiological responses to breaks in presence: A pilot study. In: Presence 2003: The 6th Annual International Workshop on Presence. Citeseer; 2003.
32. Putze S, Alexandrovsky D, Putze F, et al. Breaking the experience: Effects of questionnaires in VR user studies. In: Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems. Association for Computing Machinery; 2020:1–15. http://doi.org/10.1145/3313831.3376144
33. Slater M. How colorful was your day? Why questionnaires cannot assess presence in virtual environments. Presence. 2004,13(4):484–493. http://doi.org/10.1162/1054746041944849
34. Safikhani S, Holly M, Kainz A, et al. The influence of in-VR questionnaire design on the user experience. In: Proceedings of the 27th ACM Symposium on Virtual Reality Software and Technology (VRST '21). Association for Computing Machinery; 2021:1–8. http://doi.org/10.1145/3489849.3489884
35. Witmer BG, Singer MJ. Measuring presence in virtual environments: A presence questionnaire. Presence. 1998,7(3):225–240. http://doi.org/10.1162/105474698565686
36. Lindau A, Erbes V, Lepa S, et al. A spatial audio quality inventory (SAQI). Acta Acustica United with Acustica. 2014,100(5):984–994. http://doi.org/10.3813/AAA.918778
37. Warnecke M, Jamison S, Prepelita S, et al. HRTF personalization based on ear morphology. In: Audio Engineering Society Conference: 2022 AES International Conference on Audio for Virtual and Augmented Reality. Audio Engineering Society; 2022.
38. Møller H, Sørensen MF, Jensen CB, et al. Binaural technique: Do we need individual recordings? Journal of the Audio Engineering Society. 1996,44(6):451–464.
39. Simon LSR, Zacharov N, Katz BFG. Perceptual attributes for the comparison of head-related transfer functions. The Journal of the Acoustical Society of America. 2016,140(5):3623–3632. http://doi.org/10.1121/1.4966115
40. Hartmann WM, Wittenberg A. On the externalization of sound images. The Journal of the Acoustical Society of America.

1996,99(6):3678–3688. http://doi.org/10.1121/1.414965
41. Alais D, Burr D. The ventriloquist effect results from near-optimal bimodal integration. Current Biology. 2004,14(3):257–262. http://doi.org/10.1016/j.cub.2004.01.029
42. Blauert J. Spatial Hearing: The Psychophysics of Human Sound Localization. MIT Press; 1997. http://doi.org/10.7551/mitpress/6391.001.0001
43. Wenzel EM, Arruda M, Kistler DJ, et al. Localization using nonindividualized head-related transfer functions. The Journal of the Acoustical Society of America. 1993,94(1):111–123. http://doi.org/10.1121/1.407089
44. Iida K. Head-Related Transfer Function and Acoustic Virtual Reality. Springer Singapore; 2019. http://doi.org/10.1007/978-981-13-9745-5