Audio Engineering Society

Conference Paper

Presented at the AES 2026 International Conference on
Audio for Virtual and Augmented Reality and Immersive Games
2026 June 30 – July 3, Paris, France

This conference paper was selected based on a submitted abstract and 750-word precis that have been peer reviewed by at least two qualified anonymous reviewers. The complete manuscript was not peer reviewed. This conference paper has been reproduced from the author's advance manuscript without editing, corrections, or consideration by the Review Board. The AES takes no responsibility for the contents. This paper is available in the AES E-Library (http://www.aes.org/e-lib), all rights reserved. Reproduction of this paper, or any portion thereof, is not permitted without direct permission from the Journal of the Audio Engineering Society.

Assessing Spatial Coherence and Plausibility in a Multi-Position Virtual Choir Experience

Georgios Dedousis, Anastasia Georgaki, and Areti Andreopoulou

Laboratory of Music Acoustics and Technology, Department of Music Studies, National and Kapodistrian University of Athens, Greece

Correspondence should be addressed to Georgios Dedousis (gdedousis@music.uoa.gr)

ABSTRACT

This work reports an exploratory perceptual assessment of a binaurally rendered virtual choir, focusing on how spatial, audiovisual, and performative (including conductor-related) dimensions shape perceived plausibility of the depicted musical event. A four-part a cappella passage from Purcell's "Dido and Aeneas" was recorded at six positions, five within the choir and one at the conductor, using first-order Ambisonic recorders in a controlled acoustic environment. These recording positions were later matched to six teleportation points in a virtual performance space. The experience was reproduced over open-back headphones via a dual-engine system coupling Unity (visuals, interaction) with REAPER (audio rendering) via OSC; head-tracked scene rotation and algorithmically added late reverberation were applied in real time, with binaural decoding using an individualized KEMAR HRTF set. 24 participants with choral backgrounds wore a Meta Quest 3 headset and navigated among the teleportation points. Seven items were rated inside the virtual environment on 5-point scales: source–position connection, conductor–music temporal coherence, voice-part localizability, reverberation–room congruence, co-presence, conductor naturalness, and event plausibility. Results suggested an uneven perceptual profile where audio-spatial aspects of the scene were experienced as comparatively coherent and the depicted musical event as plausible, whereas the animated conductor was the clearest limitation. The match between reverberation and visible room size was rated as moderate, and co-presence with the choir and conductor varied widely across participants.

1 Introduction

Immersive virtual reality can make simulated events appear credible even when the user knows the experience is technologically mediated. This credibility can be decomposed into *place illusion*, the sense of being spatially situated in the virtual scene, and *plausibility illusion* (Psi), the sense that depicted events are genuinely occurring [1]. Plausibility illusion depends mainly on event credibility, contingent responsiveness, and the degree to which sensory channels cohere internally [2]. Plausibility arises also from the mutual consistency of a system's cues across sensory, perceptual, and cognitive layers, regardless of how faithfully

those cues reproduce physical reality, thus rendering it a congruence issue [3]. Both views imply that a virtual performance is judged by whether its perceptual components cohere as a credible performance situation, as well as how immersive the environment feels.

Plausibility in musical VR has been shown to be influenced by socially meaningful elements distributed across the scene. For example, an experiment on a virtual string quartet found that plausibility was influenced by the performers' gaze together with auditory factors including environmental sound, sound spatialization, and auralization [4]. A virtual rock concert suggested that salient plausibility-relevant responses came from the simulated crowd, which some participants found intrusive or disturbing [5]. Social presence in a piano-duet case study depended on embodied coregulation between performers, while a musical output from a computer-controlled agent did not sustain the subjective sense of performing together [6].

Virtual acoustic plausibility, by contrast, has focused on the idea that a binaural simulation is plausible when it matches what the listener would expect from an equivalent real acoustic scene, judged against an inner reference [7]. In VR listening over headphones, this implicates HRTF individualization/selection, and subjective quality judgments of the auditory environment depend on perceptual factors such as immersion and presence [8]. On a complementary view, immersive audio treats the listener's embodied, first-person position as its organizing principle, building meaning by exploring experiences impossible in the physical world, without needing to reproduce it faithfully [9]. For reverberation, visual room size and surface material can shape how much reverberation listeners expect to hear [10], and listeners can select the reverberation that better suits the concert hall they are shown [11], even though the perceived amount of reverberation itself does not necessarily change when seen and heard cues disagree [12].

Among headset-based systems, early work demonstrated within-ensemble virtual positioning using higher-order Ambisonic capture and 360° video without reporting a formal perceptual evaluation [13], while VR group singing has also been used to study well-being effects [14]. The acoustic basis for choral VR has advanced through simulation of room responses at multiple singing positions and rendering of position-dependent binaural auralizations [15]. Immersive rehearsal via head-mounted display and full-body avatars has reported preliminary indications of improved co-presence relative to 2D video conferencing [16]. The role of conductor gestures in such environments has not been researched extensively. Singers become more aware of gestural signals during musical ambiguity or initial unfamiliarity with a conductor, especially during performance, when memorization frees visual attention [17], and the replication of this visual and acoustic interdependence in VR can be difficult. Conductor naturalness is treated in this work as a study-specific perceptual dimension, not framed through uncanny-valley or perceptual-mismatch frameworks that were not directly measured.

Limited prior work has examined how chorally experienced listeners evaluate a non-networked, multi-position virtual choir across multiple dimensions within the same study. Existing systems have addressed these concerns partially, with some omitting head-tracked binaural rendering [18], and others concentrating on position-dependent choral auralization and acoustic/performance assessment instead of a formal perceptual evaluation of a navigable audiovisual VR choir experience [15].

The present single-condition exploratory study provides an initial perceptual assessment of a multi-position immersive virtual choir featuring head-tracked dynamic binaural rendering from first-order Ambisonic recordings. Study-specific items address the spatial, acoustic, co-presence-related, and conductor-related aspects of the experience, aiming to provide an initial perceptual profile identifying where a multi-position virtual choir appears coherent and where design refinement is needed.

2 Methods

2.1 Participants

Twenty-four participants (12 male, 12 female; range 18–32, $M = 22.6$, $SD = 2.7$) took part in the study upon giving informed consent. Musical backgrounds included sixteen musicians with choral experience, seven amateur choristers, and one choir conductor. Choral experience ranged from 1 to 15 years ($M = 5.0$, $SD = 4.3$). Twelve participants reported prior VR experience (three systematic). One participant reported partial high-frequency hearing loss in the right ear and was retained, as no exclusion criterion had been defined a

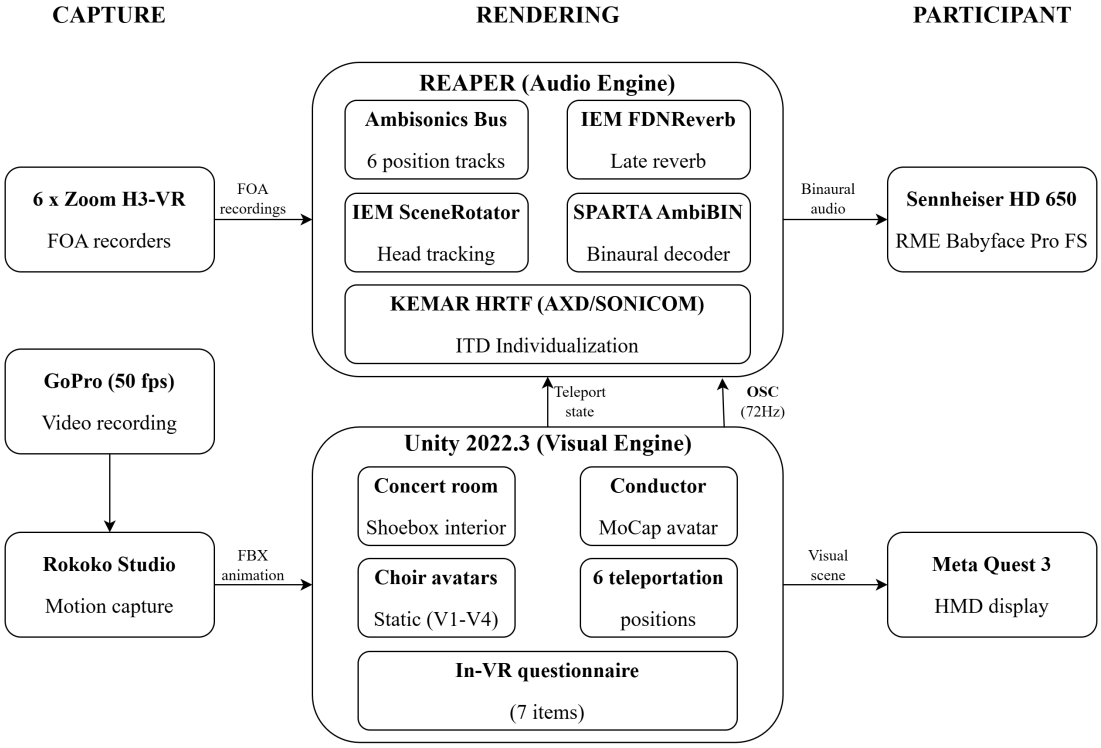


Fig. 1: System architecture: the Unity–REAPER signal chain, linked over OSC, used for visuals, interaction, and binaural audio rendering.

priori. Head circumference was measured for each participant for ITD individualization. The study received ethics approval from the Ethics and Research Deontology Committee of the National and Kapodistrian University of Athens (approval no. 259/16.9.2025).

2.2 Stimulus, virtual scene, and rendering architecture

The stimulus was a 3 min 35 s a cappella excerpt from Purcell's *Dido and Aeneas*, performed by a choir in an arc formation and recorded in a hemi-anechoic, acoustically treated space (10 m $\times$ 7 m $\times$ 5 m with absorptive wall panels, carpeted floor, and ceiling skyline diffusers) at the Laboratory of Music Acoustics and Technology (LabMAT), National and Kapodistrian University of Athens. The room's mean T30 was 0.45 s below 500 Hz and 0.29 s above 1 kHz.

Six Zoom H3-VR first-order Ambisonic (FOA) recorders captured the performance: five across the choir arc (one at each end, three between the four voice-section groups) and one above the conductor's head. These positions corresponded directly to the six teleportation spots in the virtual environment, such that each virtual listening location played back a distinct Ambisonic perspective captured from the corresponding recorder position.

Audio transport was handled in REAPER, which served as the authoritative rendering engine. Unity controlled the visual scene and interaction logic, but produced no experimental audio. The six FOA recordings were routed to an Ambisonics bus. A parallel bus containing the IEM FDNReverb plugin was summed into this signal to add artificial late reverberation, with parameters derived from the dimensions of the virtual performance space. FDNReverb's fade-in control was engaged to reduce masking of the early reflections already present in the recorded material.

Head-orientation data (yaw, pitch, roll) from the Meta Quest 3 were transmitted to REAPER via OSC, con-

figured at 72 Hz to match the device's refresh-rate mode, and drove IEM SceneRotator. The signal was decoded to binaural using SPARTA AmbiBIN. Because participant-specific HRIRs were unavailable, and to improve the binaural listening experience within the available setup, decoding used the KEMAR HRTF set from the AXD/SONICOM database [19, 20], with ITD individualization based on each participant's head circumference. This KEMAR configuration features large pinnae and GRAS ear-simulator microphones. Audio was delivered over Sennheiser HD 650 open-back headphones driven by an RME Babyface Pro FS audio interface. The rendering architecture is depicted in Figure 1.

The visual scene, rendered in Unity 2022.3.62f3 LTS and displayed via Meta Quest 3, depicted a shoebox concert-room interior with static mannequin-style choir avatars without facial features and an animated conductor avatar. The conductor motion was derived from a GoPro video captured simultaneously with the Ambisonic recordings during the same performance. The video was recorded at 50 fps and 1920×1080 resolution, then processed through Rokoko Studio's single-camera markerless motion-capture pipeline to produce a skeletal animation exported as an FBX file and imported into Unity. The resulting animation was temporally aligned with the audio using a hand-clap synchronization point present in both the video and audio recordings.

Participants teleported among the six fixed listening positions as depicted in Figure 2. Perspective changes followed a controlled transition sequence of 0.50 s fade-out, 0.20 s black hold, and 0.50 s fade-in. During the black hold, the outgoing position's recording was muted and the target position's recording unmuted, so that the discrete switch between Ambisonic perspectives fell within the visual transition and successive recordings did not overlap.

2.3 Procedure, measures, and analysis

Before the main session, participants practiced teleportation and opening, answering, and dismissing the in-VR questionnaire in a separate familiarization scene. During the main session, they listened to the excerpt while freely exploring the six listening positions, and responded to the questionnaire via the controller at any time, but were instructed to visit every position at least once during the experience.

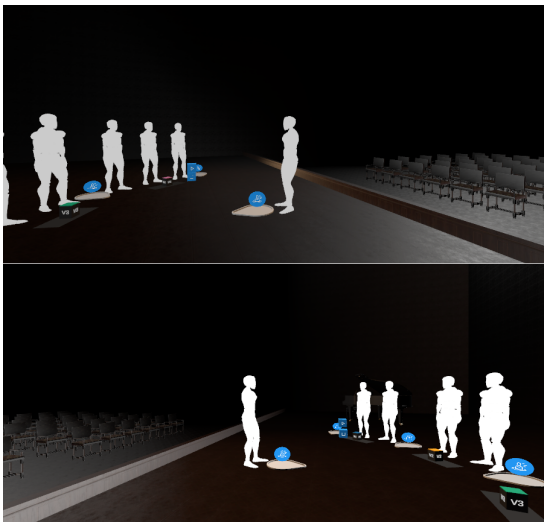


Fig. 2: Two views of the virtual shoebox performance space rendered in Unity.

Seven mandatory 5-point items were administered in Greek inside the virtual environment via a world-space questionnaire interface (English labels and anchors reported here are translations). Five items used unipolar intensity anchors (1 = Not at all, 3 = Moderate, 5 = Very much so): source–position connection, conductor–music temporal coherence, reverberation–room congruence (defined here as the perceived match between the heard reverberation and the visual size of the room), event plausibility, and co-presence with the depicted choir and conductor. On these scales, the midpoint denotes a moderate degree of the assessed quality and not neutrality. Two items used bipolar scales with a neutral midpoint: voice-part localizability (1 = Very difficult, 3 = Neither difficult nor easy, 5 = Very easy) and conductor naturalness (1 = Fully mechanical, 3 = Neutral, 5 = Fully natural). Ratings reflected the overall experience across all visited positions and were not collected per listening position.

Before the VR assessment, participants completed a paper form with demographics and background questions (age, gender, musical role, years of choral or conductor experience, self-reported hearing and vision status, and prior VR exposure). Upon completion of the assessment, they removed their headset and completed a second paper form with three open-ended prompts (two on aspects that drew them out of or into the experience, and one on possible changes that could im-

Table 1: Descriptive statistics and midpoint-check results for the seven in-VR items ($N = 24$). The midpoint column indicates the scale-center meaning: *moderate* for unipolar items and *neutral* for bipolar items. $Q1$–$Q3$ gives the interquartile interval. $n_{>3}/n_{=3}/n_{<3}$ gives participant counts above, at, and below the scale reference category (3). Raw p-values are from one-sample Wilcoxon signed-rank tests against category 3 with continuity correction; asterisks mark items surviving Holm–Bonferroni correction ($\alpha = .05$). r_{rb}: matched-pairs rank-biserial correlation.

Item	Midpoint	Mdn	$Q1$–$Q3$	$n_{>3}/n_{=3}/n_{<3}$	p	r_{rb}
Source–position connection	moderate	4	3.75–4	18 / 4 / 2	.002*	+0.76
Conductor–music temporal coherence	moderate	3.5	3–4	12 / 8 / 4	.042	+0.53
Voice–part localizability	neutral	4	4–5	20 / 4 / 0	<.001*	+1.00
Reverberation–room congruence	moderate	3	3–4	10 / 9 / 5	.079	+0.50
Event plausibility	moderate	4	3–4	14 / 7 / 3	.007*	+0.69
Co-presence	moderate	4	3–5	14 / 5 / 5	.047	+0.51
Conductor naturalness	neutral	2	2–2	1 / 4 / 19	<.001*	−0.92

prove plausibility) and a 5-point comfort rating for the VR equipment. The present paper reports only the structured in-VR ratings; the open-ended responses and equipment-comfort rating are reserved for separate analysis.

For the structured ratings, medians, IQRs, and response distributions were computed for all seven items. No composite score was computed because the items used different scale logics. Supplementary one-sample Wilcoxon signed-rank tests against the scale midpoint (3) characterized response consistency (zero differences removed before ranking, normal approximation with continuity correction). Holm–Bonferroni step-down correction across the seven tests controlled the family-wise error rate ($\alpha = .05$), and effect sizes are reported as matched-pairs rank-biserial correlations, $r_{\text{rb}} = (W^+ - W^-)/(W^+ + W^-)$. For the bipolar items, departure from the midpoint indicates a directional tendency relative to neutrality, while for the unipolar items, it indicates an evaluation above or below a moderate level, rather than presence or absence of the assessed quality. Because the design was exploratory and the items conceptually related, these tests are descriptive response-consistency checks, not confirmatory hypothesis tests. Analyses used jamovi 2.7 based on R 4.5.

3 Results

Responses across the seven in-VR items were collected from all 24 participants. No data were excluded. Descriptive and inferential statistics for all items are summarized in Table 1 and response distributions are shown in Figure 3.

The two bipolar items produced the clearest and most contrasting patterns. Voice-part localizability ($Mdn = 4$, IQR = 4–5) was the highest-rated item, with 20 of 24 participants rating it toward the "easy" end (scores: 4 or 5), and none toward the "difficult" end (scores: 1 or 2), a consistent departure above the neutral midpoint that survived Holm correction (Table 1). This result reflects the perceived ease of spatially distinguishing among voice parts. In contrast, conductor naturalness ($Mdn = 2$, IQR = 2–2) was the lowest-rated item, with 19 of 24 participants rating it toward the "mechanical" end (scores: 1 or 2). The below-neutral departure similarly survived correction.

Among the unipolar items, source–position connection ($Mdn = 4$) and event plausibility ($Mdn = 4$) were both rated above the midpoint labeled "moderate" and both survived Holm correction, indicating that participants tended to rate the spatial correspondence between visual and auditory elements, and the credibility of the depicted musical event, above a moderate level. Two participants rated source–position connection below the midpoint (the lowest score was 1), accounting for the wider spread visible in Figure 3 relative to the otherwise concentrated distribution.

Three items did not survive the Holm–Bonferroni correction. Conductor–music temporal coherence had a median just above the midpoint ($Mdn = 3.5$), but responses were distributed across the range. Co-presence ($Mdn = 4$, IQR = 3–5) was descriptively above the midpoint but heterogeneous. While a majority rated it positively (scores: 4 or 5), the remaining participants split between the midpoint and the low end, indicating that the sense of sharing a space with the virtual ensemble

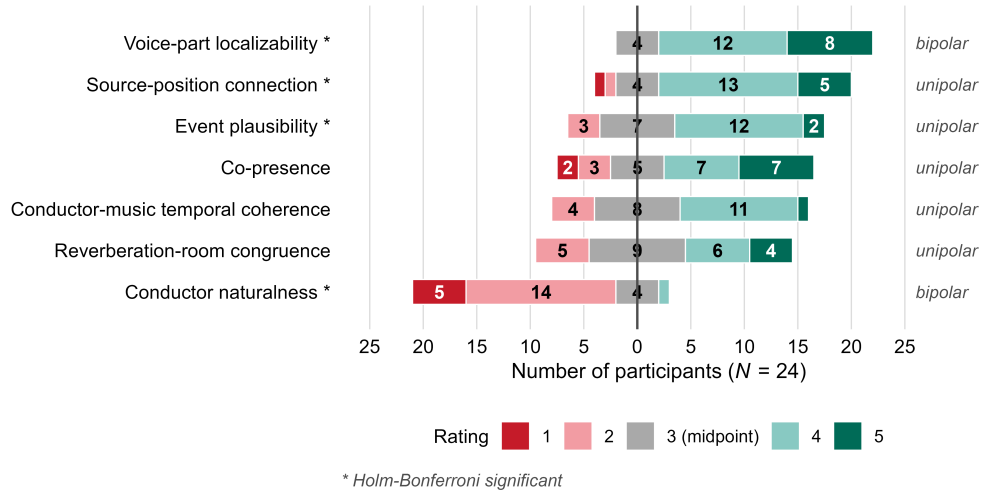


Fig. 3: Distribution of participant ratings across the seven in-VR items ($N = 24$).

varied considerably across individuals. Reverberation–room congruence ($Mdn = 3$) was the only unipolar item centered on the midpoint, indicating that participants tended to rate it as moderate.

4 Discussion

This exploratory study set out to characterize how a multi-position virtual choir was perceived across its spatial, acoustic, co-presence-related, and conductor-related aspects, and to identify its perceptual strengths and weak points. The resulting profile was uneven. The spatial organization of the choir was evaluated relatively favorably, with voice-part localizability and source–position connection rated above their reference points, and event plausibility was also rated above the moderate midpoint. By contrast, conductor naturalness was the most consistently below-neutral item, identifying the conductor animation as the clearest weakness. The reverberation item was centered on the moderate anchor, while co-presence showed wider inter-participant variation. The positive event-plausibility rating is not evidence that the spatial dimensions compensated for the conductor limitation, but it is compatible with the global plausibility rating not being determined solely by the weakest element.

Prior musical-VR studies link plausibility to the coherence of the depicted performance as a whole, including how the performers behaved and how convincing the surrounding scene was [4, 5]. Following the congruence account of [3], the present pattern indicates a partial match between the scene and participants' expectations of a choral performance across sensory, perceptual, and cognitive levels. Here, participants' ratings suggest that the perceived spatial organization of the choir corresponded relatively well to its visible layout, whereas the conductor animation was not perceived as natural. Within this non-reciprocal musical scene, the event-plausibility result should be interpreted mainly in terms of perceived audiovisual and event coherence, rather than performer responsiveness to the participant.

The conductor-related results warrant a design-specific interpretation. Low naturalness ratings may reflect the implemented motion pipeline, the markerless capture from a single camera, retargeted to a featureless mannequin avatar that could not respond to participant behavior. They should not be considered a general property of virtual conductors. Conductor–music temporal coherence was descriptively above the midpoint but did not survive correction, while naturalness stayed consistently below neutral, suggesting that these may be separable dimensions, although this claim was not formally tested.

Co-presence was descriptively positive but did not survive correction and showed the widest interquartile

range across the seven items. This variability may relate to participants being placed among pre-recorded singers and a pre-animated conductor, so the item may have captured a range of interpretations, from a scene-based impression of being among the ensemble to a stricter expectation of reciprocal musical interaction. Accordingly, the ratings may primarily reflect scene-based co-location instead of reciprocal musical co-presence, consistent with prior work emphasizing embodied co-regulation in shared VR musical performance [6]. Given the wording of the reverberation–room congruence item, its rating should be interpreted specifically as a judgment of whether the duration or amount of the added reverberation matched the visible size of the virtual room.

The single-condition design precludes causal attribution to individual system components. The seven study-specific items were conceptually related to spatial-audio and plausibility constructs but were not adopted from, or validated against, established inventories. Ratings were collected globally across all six listening positions, so it is not possible to determine whether particular vantage points shaped the overall evaluations. Because the voice-part localizability item was rated within an audiovisual scene, high scores may have reflected auditory cues together with visual section layout and source–position connection, so the item should not be read as objectively measured auditory localization performance. With $N = 24$ and no a priori power analysis, non-significant results after Holm–Bonferroni correction do not warrant the conclusion that the corresponding perceptual dimensions were unaffected.

A forthcoming extended study will report a detailed thematic analysis of the open-ended responses. The perceptual weak points identified in this study also suggest where design refinement should focus next, especially a higher-fidelity conductor representation and more systematic evaluation of reverberation–room matching. Controlled, multi-condition designs could further isolate the contributions of individual system components to these perceptual effects.

5 Acknowledgments

The research work was supported by SONICOM (www.sonicom.eu), a project that has received funding from the European Union's Horizon 2020 research and innovation program under grant agreement No. 101017743.

The research project is implemented in the framework of H.F.R.I call "3rd Call for H.F.R.I.'s Research Projects to Support Faculty Members & Researchers" (H.F.R.I. Project Number: 23677).

References

[1] Slater, M., "Place Illusion and Plausibility Can Lead to Realistic Behaviour in Immersive Virtual Environments," *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1535), pp. 3549–3557, 2009, ISSN 0962-8436, doi:10.1098/rstb.2009.0138.

[2] Slater, M., Banakou, D., Beacco, A., Gallego, J., Macia-Varela, F., and Oliva, R., "A Separate Reality: An Update on Place Illusion and Plausibility in Virtual Reality," *Frontiers in Virtual Reality*, 3, 2022, ISSN 2673-4192, doi:10.3389/frvir.2022.914392.

[3] Latoschik, M. E. and Wienrich, C., "Congruence and Plausibility, Not Presence: Pivotal Conditions for XR Experiences and Effects, a Novel Approach," *Frontiers in Virtual Reality*, 3, p. 694433, 2022, ISSN 2673-4192, doi:10.3389/frvir.2022.694433.

[4] Bergström, I., Azevedo, S., Papiotis, P., Saldanha, N., and Slater, M., "The Plausibility of a String Quartet Performance in Virtual Reality," *IEEE Transactions on Visualization and Computer Graphics*, 23(4), pp. 1352–1359, 2017, ISSN 1941-0506, doi:10.1109/TVCG.2017.2657138.

[5] Beacco, A., Oliva, R., Cabreira, C., Gallego, J., and Slater, M., "Disturbance and Plausibility in a Virtual Rock Concert: A Pilot Study," in *2021 IEEE Virtual Reality and 3D User Interfaces (VR)*, pp. 538–545, 2021, ISSN 2642-5254, doi:10.1109/VR50410.2021.00078.

[6] Van Kerrebroeck, B., Caruso, G., and Maes, P.-J., "A Methodological Framework for Assessing Social Presence in Music Interactions in Virtual Reality," *Frontiers in Psychology*, 12, 2021, ISSN 1664-1078, doi:10.3389/fpsyg.2021.663725.

[7] Lindau, A. and Weinzierl, S., "Assessing the Plausibility of Virtual Acoustic Environments," *Acta Acustica united with Acustica*, 98(5), pp. 804–810, 2012, ISSN 16101928, doi:10.3813/AAA.918562.

[8] Jenny, C. and Reuter, C., "Can I Trust My Ears in VR? Literature Review of Head-Related Transfer Functions and Valuation Methods with Descriptive Attributes in Virtual Reality," *International Journal of Virtual Reality*, 21(2), pp. 29–43, 2021, ISSN 2727-9979, doi:10.20870/IJVR.2021.21.2.4831.

[9] Geronazzo, M. and Serafin, S., "Sonic Interactions in Virtual Environments: The Egocentric Audio Perspective of the Digital Twin," in M. Geronazzo and S. Serafin, editors, *Sonic Interactions in Virtual Environments*, pp. 3–45, Springer International Publishing, Cham, 2023, ISBN 978-3-031-04020-7 978-3-031-04021-4, doi:10.1007/978-3-031-04021-4_1.

[10] Burnett, B., Neidhardt, A., Cvetković, Z., Hacıhabiboğlu, H., and De Sena, E., "User Expectation of Room Acoustic Parameters in Virtual Reality Environments," in *2023 Immersive and 3D Audio: From Architecture to Automotive (I3DA)*, pp. 1–10, 2023, doi:10.1109/I3DA57090.2023.10289314.

[11] Rolkowski, P., Odya, P., and Mróz, B., "Reverberation divergence in VR applications," *International Journal of Electronics and Telecommunications*, pp. 373–380, 2024, ISSN 2300-1933, doi:10.24425/ijet.2024.149555.

[12] Schutte, M., Ewert, S. D., and Wiegrebe, L., "The Percept of Reverberation Is Not Affected by Visual Room Impression in Virtual Environments," *The Journal of the Acoustical Society of America*, 145(3), pp. EL229–EL235, 2019, ISSN 0001-4966, doi:10.1121/1.5093642.

[13] Kearney, G., Daffern, H., Thresh, L., Omodudu, H., Armstrong, C., and Brereton, Jude., "Design of an Interactive Virtual Reality System for Ensemble Singing," in *Interactive Audio Systems Symposium, York, UK*, 2016.

[14] Daffern, H., Camlin, D. A., Egermann, H., Gully, A. J., Kearney, G., Neale, C., and Rees-Jones, J., "Exploring the Potential of Virtual Reality Technology to Investigate the Health and Well Being Benefits of Group Singing," *International Journal of Performance Arts and Digital Media*, 15(1), pp. 1–22, 2019, ISSN 1479-4713, doi:10.1080/14794713.2018.1558807.

[15] Mullins, S. S. and Katz, B. F. G., "Immersive Auralisations for Choral Ensembles," in *Proceedings on Auditorium Acoustics*, Institute of Acoustics, Athens, Greece, 2023.

[16] Di, T., Medeiros, D., Sousa, M., and Grossman, T., "VRChoir: Exploring Remote Choir Rehearsals via Virtual Reality," in *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 895–896, 2023, doi:10.1109/VRW58643.2023.00290.

[17] Jansson, D., Haugland Balsnes, A., and Durrant, C., "The Gesture Enigma: Reconciling the Prominence and Insignificance of Choral Conductor Gestures," *Research Studies in Music Education*, 44(3), pp. 509–526, 2022, ISSN 1321-103X, doi:10.1177/1321103X211031778.

[18] Dedousis, G., Andreopoulou, A., and Georgaki, A., "The Impact of Room Acoustics on Choristers' Performance: From Rehearsal Space to Concert Hall," in *Proceedings of the Stockholm Music Acoustics Conference 2023*, TRITA-EECS-RP ; 2024:4, pp. 152–158, KTH Royal Institute of Technology, Stockholm, Sweden, 2023, ISBN 978-91-8040-865-3 (print), doi:10.30746/978-91-8040-865-3.

[19] Engel, I., Daugintis, R., Vicente, T., Hogg, A. O. T., Pauwels, J., Tournier, A. J., and Picinali, L., "The SONICOM HRTF Dataset," *Journal of the Audio Engineering Society*, 71(5), pp. 241–253, 2023, doi:10.17743/jaes.2022.0066.

[20] Poole, K. C., Meyer, J., Martin, V., Daugintis, R., Marggraf-Turley, N., Webb, J., Pirard, L., La Magna, N., Turvey, O., and Picinali, L., "The Extended SONICOM HRTF Dataset and Spatial Audio Metrics Toolbox," in *Proceedings of the 11th Convention of the European Acoustics Association, Forum Acusticum 2025*, pp. 6483–6486, Málaga, Spain, 2025, doi:10.61782/fa.2025.0864.