



FORUM ACUSTICUM EURONOISE 2025

VR-PTOLEMAIC: A VIRTUAL ENVIRONMENT FOR THE PERCEPTUAL TESTING OF SPATIAL AUDIO ALGORITHMS

Paolo Ostan^{1*}

Francesca Del Gaudio¹

Federico Miotello¹

Mirco Pezzoli¹

Fabio Antonacci¹

Dipartimento di Elettronica, Informazione e Bioingegneria - Politecnico di Milano

Via Ponzio 34/5, 20133 Milano, Italia

ABSTRACT

The perceptual evaluation of spatial audio algorithms is an important step in the development of immersive audio applications, as it ensures that synthesized sound fields meet quality standards in terms of listening experience, spatial perception and auditory realism. To support these evaluations, virtual reality can offer a powerful platform by providing immersive and interactive testing environments. In this paper, we present VR-PTOLEMAIC, a virtual reality evaluation system designed for assessing spatial audio algorithms. The system implements the MUSHRA (MULTi-Stimulus test with Hidden Reference and Anchor) evaluation methodology into a virtual environment. In particular, users can position themselves in each of the 25 simulated listening positions of a virtually recreated seminar room and evaluate simulated acoustic responses with respect to the actually recorded second-order ambisonic room impulse responses, all convolved with various source signals. We evaluated the usability of the proposed framework through an extensive testing campaign in which assessors were asked to compare the reconstruction capabilities of various sound field reconstruction algorithms. Results show that the VR platform effectively supports the assessment of spatial audio algorithms, with generally positive feedback on user experience and immersivity.

Keywords: virtual acoustics, MUSHRA, virtual reality, perceptual evaluation, spatial audio

*Corresponding author: paolo.ostan@polimi.it.

Copyright: ©2025 Paolo Ostan. This is an open-access article distributed under the terms of the Creative Commons Attribution 3.0 Unported License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

1. INTRODUCTION

Spatial audio is a trending research field focused on understanding, recreating, and optimizing 3D auditory environments [1], providing listeners with an immersive and interactive listening experience. Applications range from virtual and augmented reality [2] to gaming [3], teleconferencing and remote concerts [4], where precise sound localization and an enhanced sense of presence enrich user engagement and perception.

The usual spatial audio pipeline [5,6] begins with capturing the sound scene, followed by a processing stage where the spatial information is either modified or inferred (especially for aspects that cannot be measured directly) and ends with a reproduction phase in which the (potentially altered) scene is auralized. Various specialized techniques each contribute to different aspects of this general pipeline. Among the others, we can cite for example: microphone array processing [7], which leverages strategically arranged microphones to capture audio with higher spatial resolution; Sound Field Reconstruction (SFR) [8–13], to recover the entire acoustic environment from limited microphone signals; directional source analysis [14], which studies the spatial energy distribution of sound sources; Head-Related Transfer Function (HRTF) personalization [15], that adapts the experience to an individual's head and ear shape for a more accurate listening experience; and binaural reproduction methods [16] to deliver localized sound directly to each ear, mimicking natural listening and enhancing the sense of presence. Although objective metrics play a critical role in objectively assessing many of such techniques, they do not fully capture the details that contribute to a listener's sense of realism. In fact, in spatial audio, the goal extends be-



11th Convention of the European Acoustics Association
Málaga, Spain • 22th – 26th June 2025 •





FORUM ACUSTICUM EURONOISE 2025

yond merely reproducing an accurate acoustic field; it also involves rendering a plausible and immersive experience, where listeners feel as if they are part of the scene. This subjective dimension of perception is inherently complex. Consequently, robust evaluation of spatial audio systems relies not only on objective measurement [17] but also on subjective experiments that address how listeners actually experience the audio, assessing perceptual attributes such as localization accuracy, spatial clarity, immersion, and externalization.

The standard approach for subjective sound quality evaluation is based on direct comparison methods, as outlined in the recommendations of the International Telecommunications Union (ITU) [18–20]. In recent years, methodologies such as the well-known Multi-Stimulus test with Hidden Reference and Anchor (MUSHRA) [21] have become widely adopted, offering a structured framework for assessing quality variations between multiple stimuli relative to a reference. This approach enables participants to systematically compare different versions of an audio signal, facilitating a detailed evaluation of perceived differences across specific attributes. Beyond direct comparison, other evaluation methodologies provide complementary insights into listeners' perception and evaluation of sound quality. These include sensory analysis protocols [22], indirect comparison methods [23], and behavioural analysis techniques [24].

The need of evaluation techniques for spatial audio algorithms has led to adaptations of traditional subjective testing methodologies, resulting in the development of specialized test setups. Free-field listening tests in controlled environments [25] and binaural playback evaluations using HRTFs [26] provide valuable insights into key spatial audio attributes such as localization accuracy and spatial perception. Moreover, in the last few years, with the rise of Virtual Reality (VR) and Augmented Reality (AR), spatial audio evaluation has gained new significance. VR-based assessment methods [27] offer a controlled yet immersive environment, enabling real-time testing of spatial audio reproduction in interactive scenarios. Thanks to the introduction of a visual experience, these methods enhance the perceptual realism of the virtual test scene and allow for precise evaluation of spatial immersion under dynamic conditions, supporting the refinement of spatial audio rendering techniques. Additionally, the ability to continuously track user behaviours, such as its movement and head rotation, provides valuable data for result analysis [28], offering deeper insights

into how listeners interact with spatialized audio. Finally, VR systems also enable the implementation of diverse test modalities, further expanding the scope of spatial audio evaluation [29].

In this article, we introduce VR-PTOLEMAIC: a Virtual Environment (VE) for the perceptual testing of spatial audio algorithms. Our framework integrates a real-world acoustically sampled environment (namely, a seminar room [30]) into a VR-based evaluation system. Building on previous work on sound field synthesis for 6 Degrees of Freedom (6DoF) navigation [28, 31, 32], we render spatial audio using multiple second-order Ambisonic receivers, which listeners experience through head-tracked headphones while visually exploring the VE via a VR headset. By embedding the acoustical characteristics of a real room into a virtual setting, VR-PTOLEMAIC allows for direct comparison between measured and synthesized sound fields, enhancing the validity of subjective evaluations by exploiting a MUSHRA-like testing framework. Moreover, in addition to traditional perceptual assessments, the system also incorporates real-time behavioural tracking, providing insights into how users interact with spatialized audio and the experience. This multimodal approach offers a tool for precisely evaluating spatial audio perception under realistic conditions. To validate the proposed framework, we conducted an extensive evaluation campaign in which 15 participants assessed various SFR algorithms designed to reproduce the pressure field in the virtual seminar room. The rest of the manuscript presents the evaluation system (Sec. 2), its validation (Sec. 3) and a discussion on the results (Sec. 4).

2. EVALUATION SYSTEM

Perceptual tests for spatial audio in VR require a multimodal environment that involves visual content, audio content and interactivity in order to maintain in the assessor a sense of presence and immersivity during the evaluation procedure. The following section presents a VR-based evaluation system for assessing spatial audio perception. As shown in Fig. 1, the platform integrates a VR application implemented in Unity¹ and a Max²-based audio processor, communicating via Open Sound Control (OSC) over User Datagram Protocol (UDP) for real-time interaction. By means of a VR headset with its controllers and headphones, users are able to navigate a 3D model of

¹ <https://unity.com>

² <https://cycling74.com>





FORUM ACUSTICUM EURONOISE 2025

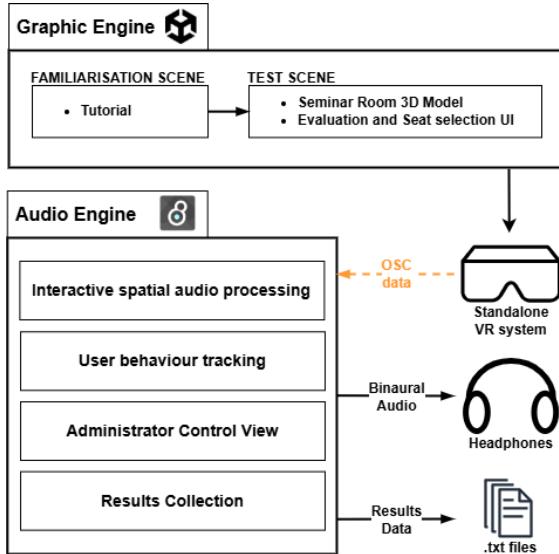


Figure 1: Overview of the proposed evaluation platform.

the test environment, select positions, and evaluate spatial audio attributes interacting with a MUSHRA-based GUI. The audio is processed dynamically allowing to listen and compare different SFRs. The stimuli proposed are the response of a convolution between anechoic sound samples and measured or reconstructed Ambisonic Room Impulse Responses (A-RIRs)s. Given that the visual content is rendered by the VR headset itself, the platform does not require a VR-ready Personal Computer (PC) in order to perform test sessions. Also, the software is designed for the Meta³ Quest 2 VR-headset, which allows for 6DoF movement tracking without the need of external sensors, and the binaural audio reproduced on closed-ear headphones does not require specific characteristics for the test room.

2.1 Visual content management

The first component of the platform is an application developed in Unity and preloaded on an all-in-one VR system. It is used to immerse the assessor in a VE, to interact with the Graphical User Interface (GUI) developed to perform the evaluations and to track user position and rotation.

The VE presented to the assessors (Fig. 2) is the 3D model of the seminar room presented in the HOMULA-



Figure 2: The 3D reconstruction of the seminar room in VR.

RIR dataset [30]. It can be noticed that, behind the main desk, 2 spheres are added to identify the location of the sound sources employed during the A-RIRs measurement campaign. The system allows for potential free navigation with 6DoF. However, due to the specific nature of the dataset [30], in which the sound field is discretized into 25 spatial positions (corresponding to the position of the chairs in the room), user movement is constrained to these predefined points. Throughout the evaluation session, the assessor can take full advantage of head rotation and exploit teleportation to move in the room by means of a dedicated UI screen, represented in Fig. 3a. The GUI for the evaluations (shown in Fig. 3b) is an implementation of the one presented in [21], with the addition of a section for source selection and a button that opens the seat selection screen. In addition, an info button located next to the attribute title shows a short description of the attribute when clicked. The UI is attached to the left controller following its movement, so it can be moved in a convenient position by the user or activated and deactivated by means of a button on the same controller.

2.2 OSC protocol setup

The interaction between the VR software and the external sound processing application is implemented through additional components taken from a Unity extension asset, extOSC⁴, that uses UDP to send customized OSC messages on an IP network. These components are C# scripts attached to all the UI elements and to the main camera object, which represents the point of view of the user. Each interaction with UI elements triggers a different OSC message according to the specific function of the element. In

⁴ <https://assetstore.unity.com/packages/tools/input-management/extosc-open-sound-control-72005>

³ <https://www.meta.com>





FORUM ACUSTICUM EURONOISE 2025

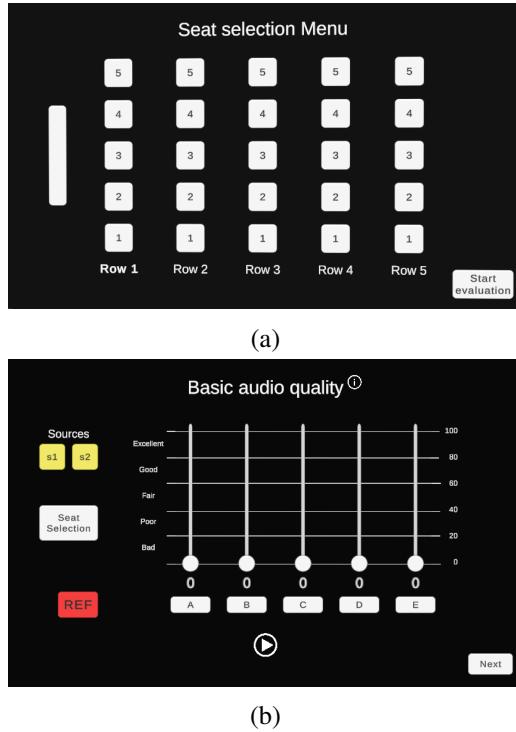


Figure 3: Seat selection UI (a) and Virtual reality implementation of the MUSHRA UI (b)

particular, buttons on the seat selection menu trigger a string type message that encodes the chosen seat position in order to recall the correct A-RIR in the sound processing phase. Also, during the entire evaluation session, the extOSC components attached to the main camera, transmit two distinct OSC messages at frame rate, encoding the camera's positional and rotational 3D coordinates.

2.3 Interactive audio content and data management

The second component of the proposed evaluation platform is a Max project running on a PC. It handles all the OSC messages received from the VR system and streams spatial audio in real-time according to user inputs on the GUI and user positional and rotational data. This component also takes care of collecting all the evaluation results and saving them in text files together with user behaviour data (time tracking, evaluated attribute and selected seat). It also serves as a control panel for the test administrator, allowing him/her to view and change settings before and during tests.

The spatial audio processing chain is reported in

Fig. 4. All the measured and reconstructed A-RIRs are imported in a dedicated buffer and recalled by the RIR selector whenever the assessor selects a spot from the seat selection menu (Fig. 3a) or switches between test items on the evaluation menu (Fig. 3b). Test administrators can select or import anechoic mono audio samples for each sound source. These are separately convolved with the related A-RIRs by the `multiconvolve~` object from HISSTools package [33] before being summed up. The resulting multichannel audio signal is sent to the SPARTA ambiBIN plugin [34] that operates real-time rotation and binaural decoding. The output of the whole pipeline is a binaural audio stream that is sent to a pair of closed-ear headphones. An optional low-pass filter can be switched on if needed, e.g. to filter the reference signal and use it as an anchor in MUSHRA [21] tests.

3. SYSTEM VALIDATION

To validate the proposed VR-based evaluation framework, we used SFR as a test scenario and conducted listening tests within the VR environment. SFR is a fundamental task in spatial audio, focused on estimating the pressure field in regions where direct measurements cannot be obtained. Capturing the acoustic field over a broad spatial area and across the entire audible frequency range requires extensive experimental resources, which is often not feasible. Instead, SFR techniques employ a limited set of observations to interpolate or extrapolate the acoustic field, offering a more practical alternative [35]. Using a MUSHRA test methodology, participants assessed various SFR methods, comparing artificially generated signals against reference signals taken from [30]. The collected responses, along with participant reliability metrics, were analyzed to assess both the accuracy of the reconstruction methods and, more importantly, the suitability of the VR-based evaluation framework in the spatial audio context.

3.1 Evaluation procedure

During the procedure, participants were first asked to complete a familiarization phase to become acquainted with the tasks, the audio samples, the controllers, the teleportation system and the UI used to evaluate reconstruction algorithms. This was followed by the assessment phase, consisting in three evaluation trials. In each trial, a different audio sample is convolved with the 25 A-RIRs generated using 4 SFR algorithms that the assessor has to evaluate according to 4 attributes with respect to an explicit reference signal. The anechoic mono audio sam-





FORUM ACUSTICUM EURONOISE 2025

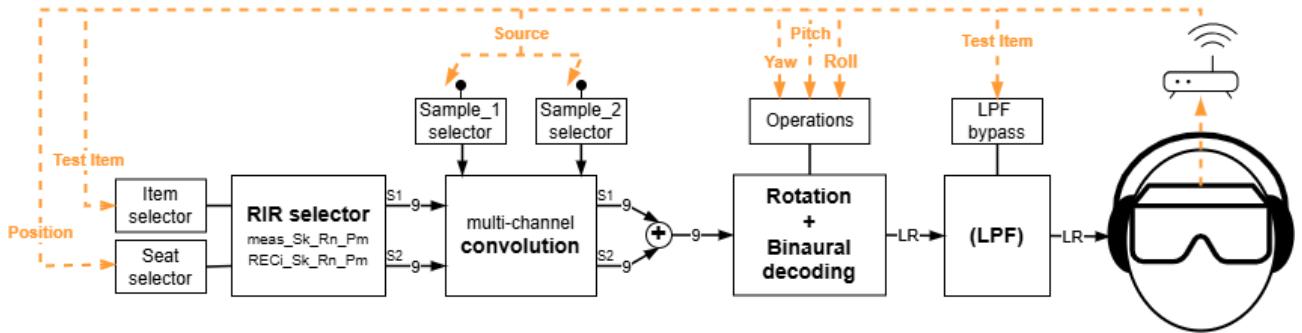


Figure 4: Interactive spatial audio processing schematics.

bles (10 to 15 s in length) were chosen from two different datasets. Specifically, the test excerpts were taken from a female voice sample and a male voice sample in the EARS dataset [36], along with an instrumental performance from the URMP dataset [37].

For the evaluation of each attribute, five different audio stimuli were proposed. Besides the explicit reference, which is the audio stream of the sample convolved with the measured A-RIRs taken from the HOMULA-RIR dataset [30], the stimuli to be evaluated were:

- Hidden reference S_{ref} - Sample convolved with the measured A-RIRs presented without explicit identification to determine evaluators reliability;
- Non-parametric solution S_A - Sound-field reconstructed employing an amplitude matching method [38];
- Low audio quality solution S_{lp} - Low-pass filtered version of the reference with a cutoff frequency of 3.5 kHz, which reduces frequency response quality while preserving spatial response quality;
- Parametric solution S_1 - Sound-field reconstructed employing a virtual mixing parametric methodology [9].

Attributes, each one characterized by a specific values range, are chosen in order to collect ratings on both sound quality and spatial audio quality:

Basic Audio Quality [0 - 100] - Global attribute used to judge all detected differences between the reference and the object [21].

Localizability [More difficult - Easier] - Attribute that correlates with the perceived spatial extent of a source. Localizability [39] is low when spatial extent and location of a sound source are difficult to estimate or appear diffuse. It is high if a sound source is clearly delimited.

Spatial Quality [Low Quality - High Quality] - A measure of the ability of the item to acoustically describe the presented scene with respect to the reference. Takes into account all spatial characteristics, e.g., depth, width, spatial distribution, reverberation, spatialization, distance, envelopment, immersion.

Timbral Quality [Low Quality - High Quality] - How accurately the item maintains the original harmonic content, tone color, and spectral balance of the sound with respect to the reference.

Test sessions were carried out on a Meta Quest 2 headset with Sennheiser HD380 Pro headphones and a Behringer MicroAMP HA400 headphone amplifier while test administration and audio processing were carried out on a laptop equipped with an Intel Core i7-7700HQ CPU and 16GB of RAM.

4. RESULTS AND DISCUSSION

The proposed VR-based system was evaluated considering its ability to facilitate subjective spatial audio assessment and its impact on user interaction. Fifteen participants (13 male, 2 female) took part in the listening tests, with an average age of 26.9 years (SD = 4.0). All but one held a university degree, and all had prior experience in music. None reported hearing impairments, while three participants had no previous experience with VR. Following the guidelines outlined in [21], four assessors were excluded from the aggregated analysis, as they rated the hidden reference condition below a specific threshold for more than 15% of the test items. The MUSHRA test results are summarized in Fig. 5, organized by the evaluated attributes. In particular, the results obtained show that participants were able to distinguish between the different reconstruction methods with a high degree of consistency.





FORUM ACUSTICUM EURONOISE 2025

Although not central to this study's validation, these findings highlight the practical efficiency, usability and utility of the proposed evaluation framework in gathering test results for spatial audio applications.

Additionally, user behaviour within the virtual scene was analyzed through head-tracking and teleportation data to gain insights into how participants interacted with the spatial environment during the evaluation. Fig. 6 presents the teleportation tracking results. In the figure, each position where a participant teleported is represented by a circle. The diameter of each circle corresponds to the amount of time spent in that location, providing an indication of where users preferred to linger during the listening tasks. Meanwhile, the greyscale shading of the circles reflects the frequency of teleportation events to that specific position: the darker the shade, the more often users moved to that spot. Participants predominantly teleported in frontal positions, likely because these areas provided clearer spatial cues and more informative perspectives for evaluating source localization and sound field characteristics. Nonetheless, they also explored the other available positions throughout the room with relative consistency.

In addition to behavioural tracking, participants completed a questionnaire at the end of the session to assess user experience and usability. Most users described the system as intuitive in terms of user interface and overall user experience, with several highlighting that the VR interface helped them better perceive differences in spatial attributes. A few users reported mild discomfort associated with wearing a head-mounted display for extended periods, primarily due to the physical weight of the device during prolonged use. These results support the system's potential as an effective platform for spatial audio evaluation, combining perceptual assessment with behavioural data and providing an informative testing environment.

5. CONCLUSION AND FUTURE WORK

We introduced a VR-based system for evaluating spatial audio reproduction, integrating real acoustic measurements into an immersive and interactive environment. The system was assessed through structured listening tests, which included a comparative evaluation of SFR algorithms, and focused on perceptual attributes such as audio quality, localizability, and spatial and timbral quality. Results demonstrated that the platform effectively supports the evaluation of spatial audio algorithms, offering perceptual feedback comparable to traditional setups. Future work will refine the user interface and tracking analysis,

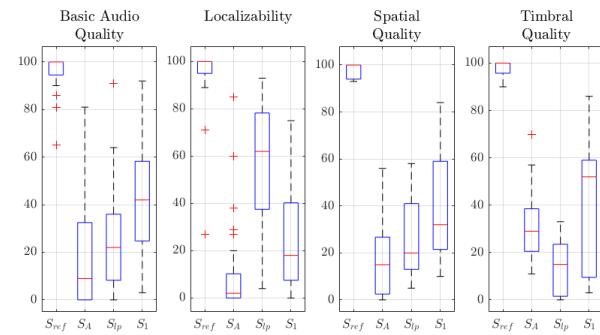


Figure 5: Results of the evaluations aggregated for attribute.

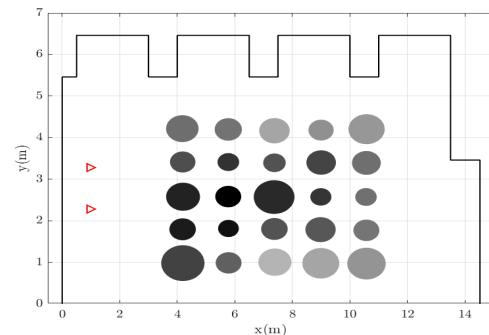


Figure 6: User behaviour tracking. Circles dimension represents the time spent in that position. The gray scale represents the number of teleportation occurrences.

with a focus on integrating detailed behavioural metrics. The simulation framework is planned for release as an open research tool.

6. ACKNOWLEDGMENTS

This work has been funded by "REPERTORIUM project. Grant agreement number 101095065. Horizon Europe. Cluster II. Culture, Creativity and Inclusive Society. Call HORIZON-CL2-2022-HERITAGE-01-02.". This work was partially supported by the European Union – Next Generation EU under the Italian National Recovery and Resilience Plan (NRRP), Mission 4, Component 2, Investment 1.3, CUP D43C22003080001, partnership on "Telecommunications of the Future" (PE00000001 – program "RESTART")





FORUM ACUSTICUM EURONOISE 2025

7. REFERENCES

[1] A. Bassuet, D. Rife, and L. Dellatorre, “Computational and optimization design in geometric acoustics,” *Building Acoustics*, vol. 21, no. 1, pp. 75–85, 2014.

[2] J. Yang, A. Barde, and M. Billinghurst, “Audio augmented reality: A systematic review of technologies, applications, and future research directions,” *Journal of the audio engineering society*, vol. 70, no. 10, pp. 788–809, 2022.

[3] H. B. Firat, L. Maffei, and M. Masullo, “3d sound spatialization with game engines: the virtual acoustics performance of a game engine and a middleware for interactive audio design,” *Virtual Reality*, vol. 26, no. 2, pp. 539–558, 2022.

[4] J. Pätynen, “A virtual symphony orchestra for studies on concert hall acoustics,” 2011.

[5] B. Rafaely, V. Tourbabin, E. Habets, Z. Ben-Hur, H. Lee, H. Gamper, L. Arbel, L. Birnie, T. Abhayapala, and P. Samarasinghe, “Spatial audio signal processing for binaural reproduction of recorded acoustic scenes—review and challenges,” *Acta Acustica*, vol. 6, p. 47, 2022.

[6] M. Cobos, J. Ahrens, K. Kowalczyk, and A. Politis, “An overview of machine learning and other data-based methods for spatial audio capture, processing, and reproduction,” vol. 2022, no. 1, p. 10.

[7] A. Politis, “Microphone array processing for parametric spatial audio techniques,” 2016.

[8] S. Damiano, F. Borra, A. Bernardini, F. Antonacci, and A. Sarti, “Soundfield reconstruction in reverberant rooms based on compressive sensing and image-source models of early reflections,” in *2021 IEEE Workshop on Applications of Signal Processing to Audio and Acoustics (WASPAA)*, pp. 366–370, 2021.

[9] M. Pezzoli, F. Borra, F. Antonacci, S. Tubaro, and A. Sarti, “A parametric approach to virtual mixing for sources of arbitrary directivity,” vol. 28, pp. 2333–2348. Conference Name: IEEE/ACM Transactions on Audio, Speech, and Language Processing.

[10] F. Miotello, L. Comanducci, M. Pezzoli, A. Bernardini, F. Antonacci, and A. Sarti, “Reconstruction of sound field through diffusion models,” in *ICASSP 2024 - 2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1476–1480. ISSN: 2379-190X.

[11] S. Damiano, F. Miotello, M. Pezzoli, A. Bernardini, F. Antonacci, A. Sarti, and T. Van Waterschoot, “A zero-shot physics-informed dictionary learning approach for sound field reconstruction,” in *ICASSP 2025-2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, IEEE, 2025.

[12] M. Olivieri, X. Karakontantis, M. Pezzoli, F. Antonacci, A. Sarti, and E. Fernandez-Grande, “Physics-informed neural network for volumetric sound field reconstruction of speech signals,” vol. 2024, no. 1, p. 42.

[13] E. Fernandez-Grande, X. Karakontantis, D. Caviedes-Nozal, and P. Gerstoft, “Generative models for sound field reconstruction,” *The Journal of the Acoustical Society of America*, vol. 153, no. 2, pp. 1179–1190, 2023.

[14] M. Pezzoli, A. Canclini, F. Antonacci, and A. Sarti, “A comparative analysis of the directional sound radiation of historical violins,” *The Journal of the Acoustical Society of America*, vol. 152, no. 1, pp. 354–367, 2022.

[15] J. C. Albarracín Sánchez, L. Comanducci, M. Pezzoli, and F. Antonacci, “Towards hrtf personalization using denoising diffusion models,” in *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2025.

[16] Z. Ben-Hur, J. Sheaffer, and B. Rafaely, “Joint sampling theory and subjective investigation of plane-wave and spherical harmonics formulations for binaural reproduction,” *Applied Acoustics*, vol. 134, pp. 138–144, 2018.

[17] J. K. Ryu and J. Y. Jeon, “Subjective and objective evaluations of a scattered sound field in a scale model opera house,” *The Journal of the Acoustical Society of America*, vol. 124, no. 3, pp. 1538–1549, 2008.

[18] R. ITU-R, “Bs. 1116-3, “Methods for the subjective assessment of small impairments in audio systems,” *International Telecommunication Union-Radiocommunication Sector*, 2015.

[19] B. ITU-R, “1534-3, “Method for the subjective assessment of intermediate quality level of audio systems,” 2015.





FORUM ACUSTICUM EURONOISE 2025

[20] T. Itu, “Itu-t recommendation p. 800. methods for objective and subjective assessment of quality,” 1996.

[21] B. Series, “Method for the subjective assessment of intermediate quality level of audio systems,” *International Telecommunication Union Radiocommunication Assembly*, vol. 2, 2014.

[22] N. Zacharov, *Sensory evaluation of sound*. CRC Press, 2018.

[23] F. Wickelmaier, N. Umbach, K. Sergin, and S. Choisel, “Scaling sound quality using models for paired-comparison and ranking data,” in *Conference Paper, DAGA 2012 Congress 38th German Annual Conference on Acoustics*, 2012.

[24] C. Pike and H. Stenzel, “Direct and indirect listening test methods—a discussion based on audio-visual spatial coherence experiments,” in *Audio Engineering Society Convention 143*, Audio Engineering Society, 2017.

[25] C. Freigang, K. Schmiedchen, I. Nitsche, and R. Rübsamen, “Free-field study on auditory localization and discrimination performance in older adults,” vol. 232, no. 4, pp. 1157–1172.

[26] A. Andreopoulou and B. F. Katz, “On the use of subjective hrtf evaluations for creating global perceptual similarity metrics of assessors and assessees.,” in *ICAD*, pp. 13–20, 2015.

[27] E. Medina, R. Viveros-Muñoz, and F. Otundo, “Assessing ambisonics sound source localization by means of virtual reality and gamification tools,” *Applied Sciences*, vol. 14, no. 17, p. 7986, 2024.

[28] O. Rummukainen, T. Robotham, S. J. Schlecht, A. Plinge, J. Herre, and E. A. Habels, “Audio quality evaluation in virtual reality: multiple stimulus ranking with behavior tracking,” in *Audio Engineering Society Conference: 2018 AES International Conference on Audio for Virtual and Augmented Reality*, Audio Engineering Society, 2018.

[29] S. Gorzynski, N. Kaplanis, and S. Bech, “A flexible software tool for perceptual evaluation of audio material and vr environments,” in *Audio Engineering Society Convention 149*, Audio Engineering Society, 2020.

[30] F. Miotello, P. Ostan, M. Pezzoli, L. Comanducci, A. Bernardini, F. Antonacci, and A. Sarti, “HOMULA-RIR: A room impulse response dataset for teleconferencing and spatial audio applications acquired through higher-order microphones and uniform linear microphone arrays,” in *2024 IEEE International Conference on Acoustics, Speech, and Signal Processing Workshops (ICASSPW)*, pp. 795–799.

[31] L. McCormack, A. Politis, T. McKenzie, C. Hold, and V. Pulkki, “Object-based six-degrees-of-freedom rendering of sound scenes captured with multiple ambisonic receivers,” vol. 70, no. 5, pp. 355–372. Publisher: Audio Engineering Society.

[32] A. Figueroa-Duran and E. Fernandez-Grande, “Reconstruction of reverberant sound fields over large spatial domains,” *The Journal of the Acoustical Society of America*, vol. 157, no. 1, pp. 180–190, 2025.

[33] A. Harker and P. A. Tremblay, “The hisstools impulse response toolbox: Convolution for the masses,” in *Proc. of the international computer music conference*, pp. 148–155, The International Computer Music Association, 2012.

[34] L. McCormack and A. Politis, “Sparta & compass: Real-time implementations of linear and parametric spatial audio reproduction and processing methods,” in *AES International Conference on Immersive and Interactive Audio*, Audio Engineering Society, 2019.

[35] S. Damiano, F. Miotello, M. Pezzoli, A. Bernardini, F. Antonacci, A. Sarti, and T. Waterschoot, *A Zero-Shot Physics-Informed Dictionary Learning Approach for Sound Field Reconstruction*.

[36] J. Richter, Y.-C. Wu, S. Krenn, S. Welker, B. Lay, S. Watanabe, A. Richard, and T. Gerkmann, “EARS: An anechoic fullband speech dataset benchmarked for speech enhancement and dereverberation,” in *Interspeech 2024*, pp. 4873–4877, ISCA.

[37] B. Li, X. Liu, K. Dinesh, Z. Duan, and G. Sharma, “Creating a multitrack classical music performance dataset for multimodal music analysis: Challenges, insights, and applications,” vol. 21, no. 2, pp. 522–535.

[38] T. Abe, S. Koyama, N. Ueno, and H. Saruwatari, “Amplitude matching for multizone sound field control,” *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 31, pp. 656–669, 2022.

[39] A. Lindau, V. Erbes, S. Lepa, H.-J. Maempel, F. Brinkman, and S. Weinzierl, “A spatial audio quality inventory (saqi),” *Acta Acustica united with Acustica*, vol. 100, no. 5, pp. 984–994, 2014.

