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, Sorbonne University/d'Alembert & IRCAM/STMS, Paris, France

This paper was peer-reviewed as a complete manuscript for presentation at this conference. 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.

Real-Time Spatial Auralization for Collaborative Architectural Design in Virtual Reality

Achilleas Xydis[1], Giacomo Montiani[1], Chao Chia-Hsuan[1], and Fabio Scotto[1]

[1]*Gramazio Kohler Research, ETH Zurich*

Correspondence should be addressed to Achilleas Xydis (xydis@arch.ethz.ch)

ABSTRACT

Acoustic performance remains insufficiently addressed in architectural design, where visual and spatial considerations typically guide decision-making. The present research examines the integration of real-time spatial auralization into multi-user virtual reality environments to facilitate collaborative evaluation of architectural acoustic performance during design exploration. A multi-user VR framework has been developed that embeds real-time binaural auralization within a collaborative design evaluation context. The framework supports shared inhabitation of virtual models, auditory exploration of spatialized sound sources, and systematic investigation of how spatial configurations and material properties affect perceived acoustic characteristics. Temporal soundscapes and acoustic implications of alternative materials and spatial arrangements can be evaluated through a gesture-driven interface. Spatialized voice communication further supports co-inhabitation of virtual environments and acoustically informed discussion. By embedding real-time auralization within collaborative virtual environments, the framework repositions acoustic performance as an experiential dimension of architectural design, enabling earlier incorporation of auditory considerations into the design process. An exploratory user study indicates that immersive, real-time auralization can integrate acoustic feedback into collaborative architectural design workflows, supporting multisensory evaluation and reducing dependence on late-stage corrective interventions.

1 Introduction

Acoustic performance is an important aspect of architectural design, influencing speech intelligibility, privacy, comfort, and the overall perception of space. However, it is rarely considered during the early stages of architectural design, where spatial arrangement, circulation, and visual qualities typically dominate decision-making. While acoustical simulation tools can provide detailed analysis, they are usually employed after key spatial and material decisions have already been made. As a result, acoustic performance is often addressed through late-stage corrective interventions rather than being integrated as a core design parameter.

At the same time, architectural design is inherently collaborative. Architects, engineers, consultants, and clients often evaluate design proposals together, discussing spatial qualities and trade-offs during iterative design reviews. Over the past two decades, immersive virtual environments have been explored as tools for collaborative architectural design and design review. Multi-user VR platforms allow stakeholders to inhabit shared digital models at full scale

and collaboratively evaluate spatial proposals, supporting communication, participatory design, and stakeholder engagement [1, 2, 3]. Immersive environments have been shown to enhance spatial understanding and communication in design compared to traditional two-dimensional representations [4].

Despite these advances, most collaborative VR platforms for architectural design primarily emphasise visual and spatial aspects of the built environment. Acoustic properties are rarely incorporated into these systems beyond basic audio effects used for environmental immersion. Physically informed acoustic simulation and perceptual auralization remain largely confined to dedicated acoustic simulation tools or specialised research systems. As a result, acoustic evaluation is seldom integrated into immersive collaborative design workflows. Integrating perceptually meaningful acoustic simulation into collaborative virtual environments could allow design teams to evaluate auditory consequences of spatial and material decisions while they are still being developed. Instead of relying solely on numerical analysis performed after the design has largely been defined, designers could directly experience how design alternatives shape a space's acoustic character.

The research presented in this paper builds upon collaborative VR paradigms by integrating real-time acoustic simulation and binaural auralization into a shared virtual environment, enabling design teams to explore the auditory consequences of spatial and material decisions during collaborative design evaluation.

2 Related work

Auralization offers an alternative approach for communicating acoustic performance, allowing listeners to experience how sound propagates within a space [5]. Studies on professional acoustic practices indicate that auralization is frequently used to communicate acoustic results to non-experts, as auditory demonstrations are often easier to understand than numerical reports [6]. Auralization techniques have been applied to architectural acoustics, environmental noise studies, and urban soundscape analysis [7, 8]. However, traditional auralization workflows are typically based on precomputed impulse responses and fixed listening positions, limiting users' ability to explore acoustic conditions freely within a virtual environment.

2.1 Real-time collaborative evaluation

Research on interactive acoustic simulation has explored ways to enable dynamic auditory experiences. Early work on interactive virtual acoustics demonstrated the possibility of combining virtual visual environments with real-time acoustic simulation [9]. Later approaches introduced perceptually optimised rendering techniques and GPU-based simulation pipelines to support interactive auralization of complex environments [10, 11]. Canfield-Dafilou et al. used a real-time auralisation system [12] to study the interplay between composing and performing music, and room acoustics [13]. More recent research has investigated hybrid simulation approaches that update acoustic responses in real time as scene geometry or listener position changes [14]. RWTH Aachen University pursued a multi-modal system combining adaptive cross-talk-cancellation loudspeaker reproduction, infrared camera-based head and object tracking, and a low-latency, non-uniformly partitioned convolution engine capable of processing more than 50 sources in real time [15]. Despite these advances, most implementations remain focused on single-user experiences or specialised laboratory setups rather than collaborative design environments.

2.2 Collaborative virtual environments in architecture

Immersive virtual reality can support spatial understanding, design review, and participatory decision-making by allowing users to experience architectural proposals at full scale [16]. Multi-user virtual environments further extend these capabilities by enabling multiple participants to inhabit and interact within shared digital models simultaneously, supporting collaborative exploration of spatial configurations and design alternatives [17]. Such environments allow architects, engineers, and stakeholders to collectively evaluate spatial scale, circulation, and visual qualities of proposed designs within a shared immersive context.

2.3 Spatial audio and performance constraints

Spatial-audio reproduction techniques enhance the ecological validity of virtual environments by preserving inter-aural cues and source directivity [18]. Allowing designers to compose time-ordered sound events in VR

Table 1: Steam Audio Material settings.

	Material	Absorption %			Scattering %	Transmission %		
		Low	Mid	High		Low	Mid	High
0	Concrete Wall (300mm)+Plaster (10mm)	0.01	0.02	0.04	0.25	0.02	0.01	0.01
1	Concrete painted	0.01	0.02	0.02	0.25	0.02	0.01	0.01
2	Wood flooring on joists	0.13	0.08	0.06	0.25	0.02	0.01	0.01
3	Glass door (4mm glass)	0.15	0.04	0.02	0.15	0.15	0.02	0.02
4	Double glassing (3mm glass+10mm airgap)	0.04	0.03	0.02	0.15	0.07	0.03	0.05
5	Thin metal	0.08	0.02	0.01	0.15	0.2	0.02	0.01
6	Felt 50mm	0.4	0.8	0.99	0.25	0.5	0.01	0.01
7	Felt 25mm	0.25	0.95	0.99	0.25	0.7	0.05	0.01
8	Solid Timber Door	0.04	0.05	0.05	0.2	0.3	0.1	0.05
9	Wooden chair - empty	0.04	0.06	0.08	0.25	0.9	0.9	0.9
10	Person seating	0.35	0.5	0.6	0.5	0.5	0.3	0.1
11	Person standing	0.3	0.45	0.5	0.6	0.5	0.3	0.1
12	Drapery (600 gr/m2, pleated 50%)	0.25	0.65	0.65	0.35	0.35	0.15	0.1

increases perceived authenticity by up to 30% compared with static impulse-response playback [19]. Hybrid GPU beam-tracing or ray-tracing pipelines achieve interactive updates. Still, they can reach 5-10ms per source per frame, quickly exceeding the 16ms frame budget of 60 Hz for VR applications as scene complexity increases [14]. Pre-computed beam trees reduce cost, yet any material or layout change triggers multi-second rebuilds, interrupting design iteration [20]. Edvaldsson and Xydis introduced a method for latency-free real-time auralization with self-excitation [21]. This approach uses a loudspeaker array within a semi-anechoic room and leverages the room's natural floor reflections to adjust the impulse response and compensate for the total latency of the loudspeaker system.

3 Methodology

This paper presents the audio architecture of EchoXR [22], a VR framework built in Unity (Unity Technologies)[1] that embeds real-time binaural auralization directly inside a collaborative, multi-user environment. The system conveys acoustic performance through listening rather than numbers, allowing architects, consultants, and clients to co-inhabit, co-experience, and evaluate the auditory and spatial qualities of designs in real time.

The framework allows users to hear environmental sounds and each other's voices dynamically auralized based on the spatial layout, surface materials, and their position in the space. 6-DoF tracking for the head's position and orientation enables users to move freely and perceive changes in sound sources and room responses in real time. Users physically move in their real environment, and these movements are reflected in corresponding movements in the virtual space, reinforcing spatial coherence between auditory and visual cues. The auralization is synced with the system's visual and interactive elements, ensuring that any design modifications, such as changing materials or adding or removing elements, result in immediate updates to the acoustics. Furthermore, these changes occur without interrupting the immersive experience, enabling a seamless multisensory environment for collaborative reviews. This capability allows users to engage in real-time discussions, try out several design and material options, and experience how they look and how they influence the acoustics of the space. This process enables a more informed decision-making process that includes acoustics as an evaluation criterion.

The framework's audio architecture comprises several components: acoustic material modelling, sound source characterisation, soundscape design, real-time auralization, multi-user voice communication, and design options comparison. Steam Audio (Valve Corporation)[2] is used to model the acoustic properties of the space's surfaces, characterise sound sources, and gen-

[1] https://unity.com/games

[2] https://valvesoftware.github.io/steam-audio/



Fig. 1: Acoustic material assignment. A Steam Audio Material is assigned to all acoustically relevant meshes in the scene. These materials can be used across multiple meshes in the scene.

erate real-time binaural auralizations. Unity's timeline editor is used to control when a sound is present in the scene and how long it lasts. Photon Fusion (Exit Games GmbH)[3] is used to exchange data among users and to enable a gesture-driven interface for design options and material toggling. Photon Voice is used for low-latency voice communication. Finally, the Meta Quest SDK (Meta)[4] is used for immersive VR interactions, and the Meta Mixed Reality Utility Kit (MRUK) (Meta)[5] for tracking and handling the user's position and orientation. The following sections describe these components.

3.1 Acoustic materials

The Steam Audio Material component is applied to all acoustically relevant meshes (Figure 1). This component enables the characterisation of acoustical properties, such as absorption, scattering, and transmission, across low-, mid-, and high-frequency bands, enabling more accurate acoustic modelling (Table 1).

[3]https://www.photonengine.com/voice
[4]https://developers.meta.com/horizon/reference
[5]https://developers.meta.com/horizon/documentation/unity/unity-mr-utility-kit-overview/

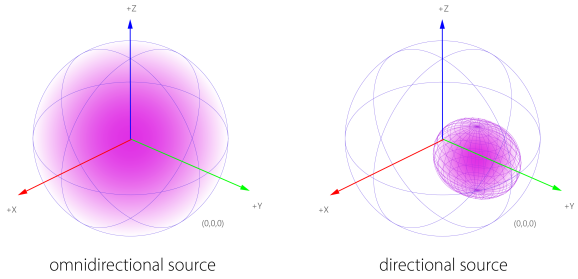


Fig. 2: Sound source's directivity pattern. Left: Omnidirectional source emitting sound in every direction. Right: Directional source emitting sound only along the 0-degree Axis.

3.2 Sound sources

Convincing sound source characterisation is crucial for realistic auralizations. Therefore, all sound sources, such as voices, equipment, appliances, environmental sound events, and more, are assigned an audio file, a volume, and a directivity pattern (Figure 2). For example, a kettle is an omnidirectional sound source that radiates sound equally in all directions. On the other hand, a human is a directional sound source, radiating more sound forward than behind. In the absence of an anechoic chamber for recording dry signals, the close-mic technique was chosen as a quick, flexible alternative. The source's real-world volume level is measured in decibels (dB) using an NTi XL3 Sound Level Meter (SLM). Unity and Steam Audio use a volume range from 0.00 to 1.00; therefore, the various sound source volume levels are scaled to approximate real-world sound levels. Finally, the source's directivity pattern is either taken from the literature or measured using the SLM around the source.

Sound sources can be stationary or movable. Stationary sources are fixed objects, such as seated people having phone calls, kitchen appliances, TVs, HVAC systems, and more, while movable sources are objects such as people walking or vehicles. Movable sound sources are Photon Fusion's Network Objects. A Network Object is a Unity `GameObject` with a `NetworkObject` component. This component transmits the location and orientation of the sound source in real-time to all users, enabling them to actively interact with scene objects, moving them or passing them back and forth as an additional medium for experiencing the space's acoustics.

3.3 Soundscape design

Immersion, the subjective impression that one participates in a realistic experience, while cut off from the world [23, 24], is also influenced by the sounds present in the virtual environment [25]. The soundscape [26] of a physical space comprises multiple static or moving, overlapping, and repeating sound events. To increase acoustic immersion and realism in the presented framework, designers can curate complex spatial and temporal soundscapes by controlling the position, orientation, starting time, and duration of sound events. Sound sources are stored as Unity prefabs, each containing settings for which sound file to play, its volume, and its directivity pattern. The position and orientation of a sound source are set by placing the appropriate prefab inside the Unity scene (Figure 3). Using Unity's timeline editor, sound sources can start and stop at a specific time, repeat as many times as needed, and move in the virtual space. Figure 4 shows the timeline editor and all the sound events in the scene. Multiple sounds can play simultaneously, but what the user hears depends on their distance from them and the scene's acoustics. The framework instantiates the prefabs at the predefined positions and scheduled times, enabling complex, context-rich soundscapes that mirror real-world scenarios. In multi-user virtual experiences, the timeline is triggered once all users have joined the virtual model, ensuring their soundscapes are synchronised.

3.4 Real-time auralization

The user's head position and orientation are tracked by the Quest 3's tracking system. These data, together with a generic Head-Related Transfer Function (HRTF) set in the Steam Audio's general settings, provide a real-time, realistic binaural auralization over headphones. HRTFs describe how a listener's head, pinnae, and torso filter incoming sound waves before they reach the inner ear. By modelling this influence, convolving signals with HRTFs synthesises realistic spatial audio, enabling externalisation and localisation of virtual sound sources. As users move within the modelled virtual space, they can experience smooth sound transitions and accurately perceive the spatial positioning of each sound source, as in a real-world scenario.

3.5 Spatialized voice chat

The presented framework extends real-time collaborations by enabling spatialized voice communication between users. When the user speaks to their microphone,

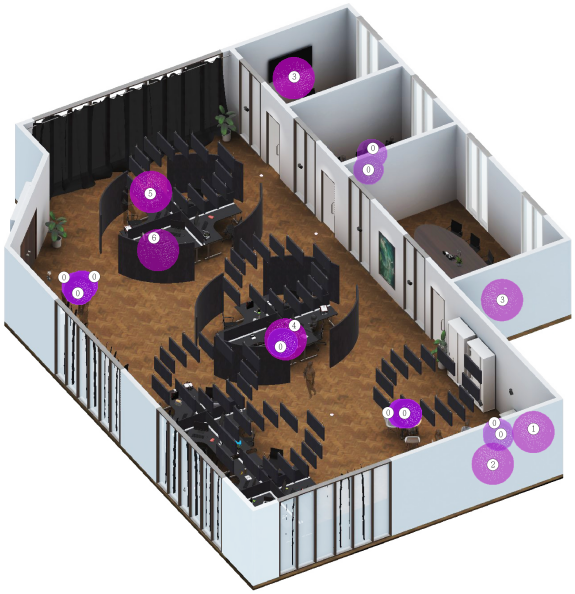


Fig. 3: Sound sources inside the scene. The pink spheroids represent the location and directivity pattern of each sound source.

Photon's Recorder component captures the audio and streams it over the Photon Unity Networking system. The audio is played back to the other user through a sound source attached to the speaking player's avatar, which moves with them in the scene. Steam Audio spatializes sound using the distance between users, the speaker position in the model, and the acoustic properties of each mesh in the scene. The result is a perceptually real-time auralized and location-accurate voice communication. Figure 5 shows how a user's voice is played through a virtual source at the location of their avatar's head (avatar A). The other user (avatar B) hears user A's voice from the right, along with the room's acoustic response.

3.6 Design-option comparison

A palm-anchored UI enables users to toggle predefined spatial configurations or change surface materials using gestures. All toggleable meshes in the scene receive the Steam Audio Dynamic Object components. Game objects with this component can update their geometry or acoustic material properties at runtime without requiring a full re-bake of the acoustic simulation, enabling rapid, seamless A/B comparisons of alternative options.

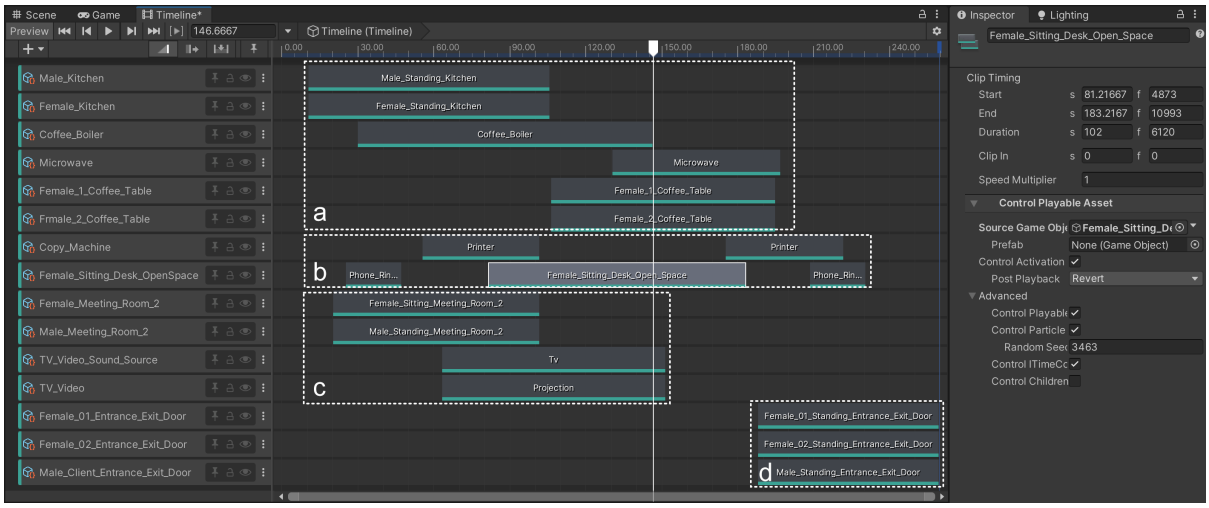


Fig. 4: The soundscape arrangement inside Unity's timeline editor. The blocks control the playback of each sound source's audio clip (Inspector panel on the right). The dotted squares show: a) sounds present in the kitchen area, b) at a desk island, c inside two meeting rooms, and d) of a group of people chatting near the entrance.

Figure 6 shows two design options for an open-plan office space. The top-left section shows three semi-open meeting rooms at the back of the space. In the bottom-left section, the user toggles the plasterboard walls and glass doors to all meeting rooms. Similarly, the top-right section shows the desk islands next to each other, and the open kitchen corner at the back of the room. In the bottom-right section, the user toggles the 1.8-meter-tall felt partitions.

4 Case-study: Collaborative acoustic design evaluation

To explore whether this approach can enable acoustic performance evaluation in architectural design, we invited 6 trained architects, 5 architecture students, 3 trained engineers, and 2 acousticians to use the framework and perform a series of tasks. The study focused on evaluating the acoustic implications of alternative spatial configurations and material choices. Participants experienced a virtual environment in pairs using Meta Quest 3 head-mounted displays and headphones. The qualitative data collection combined in-session observation of participant interactions with a post-session questionnaire. During the session, participants' navigation patterns, design toggling behaviour, and verbal exchanges were observed and documented to capture how they interacted with the system and how collaborative evaluation unfolded in practice. In the post-session questionnaire, participants assessed their perceptions of the acoustic differences, the clarity of the comparison between design options, and the overall usefulness of the system for design decision-making.

4.1 Design scenarios

The environment is an open-plan office with three office division design scenarios and four acoustic treatment solutions that can be turned on and off independently: Scenario 1 has no divisions between the meeting rooms and the main office area. Scenario 2 adds acoustic curtains around the meeting rooms, and scenario 3 substitutes the curtains with plasterboard walls and glass doors. The toggleable acoustic treatments are a) curtains in front of the façade windows, b) sound absorptive baffles on the ceiling, c) 60cm high vertical panels on each desk, and d) 1.8m high acoustic dividers between desk islands.

4.2 Collaborative design evaluation

During the evaluation session, participants were asked to freely move within the virtual environment while listening to a dynamic soundscape consisting of speech, office equipment, and environmental sounds. Using the

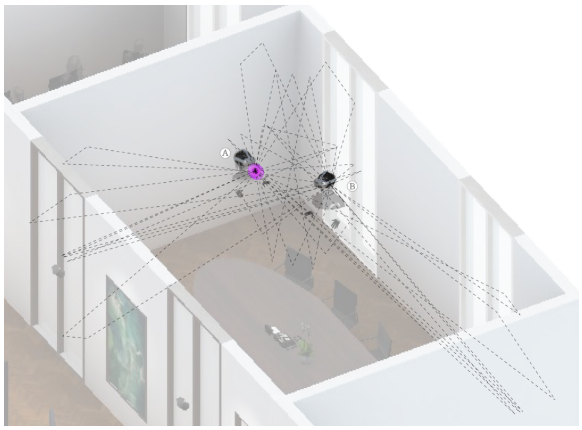


Fig. 5: User's voice spatialisation. The virtual sound source is located at the avatar's position, with the directivity pattern pointing forward. The dotted lines indicate some of the early reflections arriving at the other user's avatar.

gesture-based interface, users switched between spatial configurations and material options in real time while discussing the perceived acoustic differences.

4.3 Study results

The results are based on participants' questionnaire responses and observations of their interactions during the VR sessions and are grouped into three topics:

Perception and evaluation: Participants reported audible differences across the tested spatial configurations. In the open-plan scenario, the perceived acoustic difference was highest for ceiling panels ($M = 3.00, SD = 0.00$) and island panels (M = 2.69, SD = 0.48), compared with desk panels (M = 1.75, SD = 0.45) and window curtains (M = 1.00, SD = 0.00). A similar pattern was observed for the plaster-walls-and-glass-doors configuration, where ceiling panels (M = 3.00, SD = 0.00) and island panels (M = 2.69, SD = 0.48) were again rated as more noticeable than desk panels (M = 1.00, SD = 0.00) and window curtains (M = 1.00, SD = 0.00). In the acoustic-curtains scenario, ceiling panels remained the most noticeable intervention (M = 2.38, SD = 0.50), while desk panels and window curtains were again rated low (both M = 1.00, SD = 0.00).

Collaboration: In-session observations showed that participants actively used the tool to collaboratively compare alternatives. They switched between spatial configurations and acoustic treatments, moved to different listening positions, and discussed the resulting acoustical changes. Communicating with one another in the system was rated highly (M = 4.63, SD = 0.50), hearing the other participant's voice auralized in the simulated room was rated highly (M = 4.69, SD = 0.48), and cohabiting the same virtual environment received a mean rating of 4.75 (SD = 0.45).

Usefulness: Participants found the auralization helpful in understanding how spatial layouts and material choices influence acoustic qualities such as speech clarity, background noise levels, perceived reverberation, and sense of privacy. The highest usefulness rating was given to the ability to explore different design options and instantly hear their impact (M = 4.94, SD = 0.25). Freely walking within the virtual space was also rated positively (M = 4.19, SD = 0.40). Several participants noted that experiencing these effects directly and in real time made it easier to evaluate trade-offs than relying on numerical acoustic descriptors. Finally, the spatialized voice communication enabled natural dialogue, allowing participants to point out acoustic effects, discuss alternatives, and reach shared conclusions. This process shifts the evaluation from an individual perspective to a collective understanding.

5 Conclusions and future work

This paper presents a real-time spatial auralization framework that embeds perceptual acoustic feedback into collaborative architectural design sessions. By combining binaural rendering, dynamic material and geometry manipulation, and spatialized voice communication, the system enables multiple participants to inhabit and evaluate design options in a shared virtual environment. The integration of frequency-band material modelling, directivity-aware sound-source calibration, and synchronised user interaction establishes a workflow in which acoustics becomes an immediate, embodied parameter during the design phase. The ability to toggle between material configurations and spatial arrangements, while hearing their acoustic consequences in real time, facilitates multisensory decision-making and cross-disciplinary dialogue. The framework shifts acoustic evaluation from post-design numerical analysis toward a perceptual and collaborative design activity.

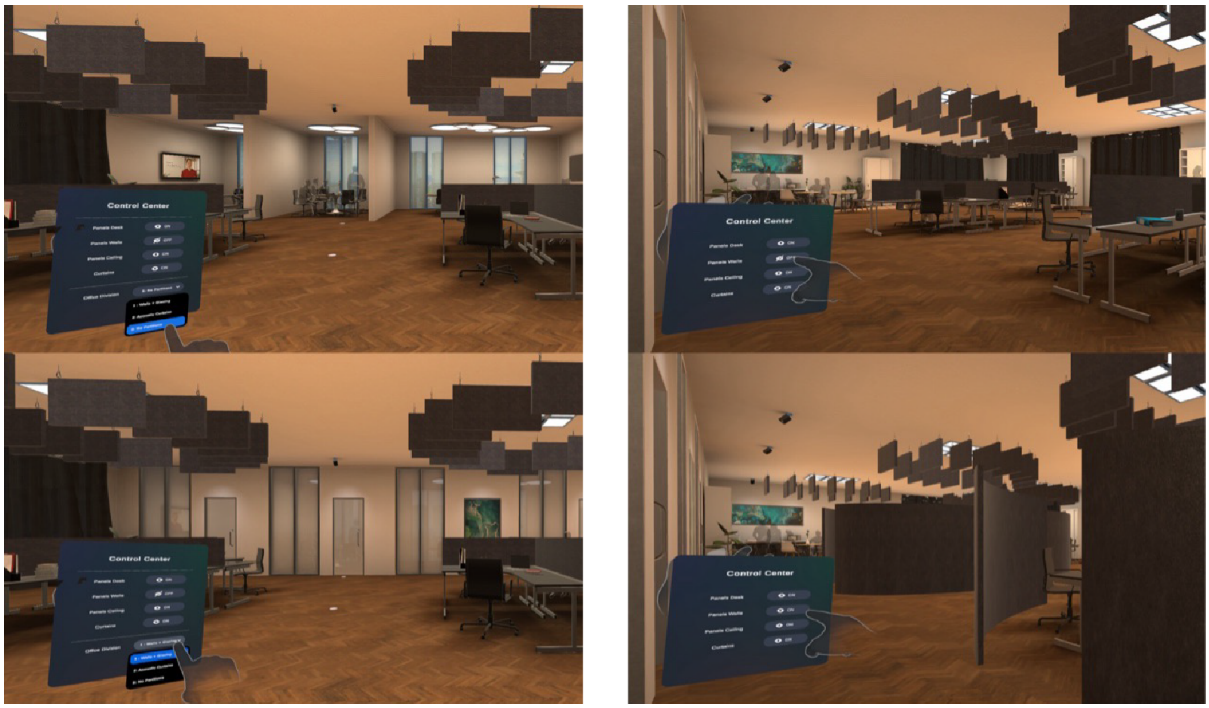


Fig. 6: Using the palm-anchored UI, users compare different design options. Left: toggling meeting room walls. Right: toggling of acoustic partitions.

By enabling architects and consultants to jointly experience the auditory consequences of design decisions, acoustic performance can be considered earlier in the design process. Traditional room-acoustics tools provide accurate acoustic simulations but typically rely on offline analysis and numerical interpretation. In contrast, the proposed framework prioritises perceptual evaluation through real-time auralization, allowing design teams to explore acoustic consequences interactively during collaborative design sessions.

The current implementation demonstrates the feasibility and value of bringing real-time auralization into immersive design reviews, yet limitations persist. Acoustics simulation precision is an important factor in achieving representative acoustic conditions. The framework uses the Steam Audio plug-in for room-acoustics simulation and auralization. The plug-in models distance attenuation, occlusion, scattering, and source directivity pattern. It also enables control over pre-computed and real-time sound reflections at design time, allowing the optimisation between quality and performance. Yet, it is a plug-in designed for computer games, not for accurate room-acoustic simulations. A comparison with state-of-the-art Geometrical Acoustics simulations can help evaluate its validity and identify its limitations.

Unity's audio engine imposes a minimum round-trip delay of around 100ms between microphone sampling (`Microphone.Start`) and local loudspeaker playback [27]. This delay is significantly higher than the around 30ms perceptual threshold at which people begin to detect echoes or rhythmic disruption when monitoring their own voice [28]. Therefore, real-time convolution of a participant's own speech would introduce objectionable comb filtering and temporal smear. Furthermore, mixing auralized and non-auralized sources generally reduces immersion in extended reality applications. Therefore, self-voice auralization is omitted, and the local user hears only the unprocessed bone-conducted signal. At the same time, the captured stream is transmitted to remote peers, who receive a fully spatialized rendering with no perceptual latency.

Voice communication is implemented using Photon Voice 1. The voice streams of all users are combined

into a one audio channel and streamed to all users. This limits the number of users who can be properly auralized within the scene. Nevertheless, Photon Voice 2 supports discrete audio streams for each user, enabling voice spatialisation for larger groups. Furthermore, a curated library of acoustically informed prefabs bundling geometry, textures, acoustic materials, and sound sources, along with an extended acoustic materials library, can reduce authoring overhead and improve the fidelity and expressiveness of auralizations. Lastly, even though custom HRTFs are supported in the current implementation, they can only be assigned before the app is built. Implementing dynamic selection of HRTF files at runtime can enable more accurate, personalised binaural rendering.

Acknowledgements

This research was funded by the ETH Foundation. The authors thank the Center for Augmented Computational Design in Architecture, Engineering and Construction (Design++) at ETH Zurich for providing access to the Immersive Design Lab for developing and testing the EchoXR research project.

Contributions

Fabio Scotto (FS) and Achilleas Xydis (AX) jointly conceived the project's original idea and supervised the overall research. FS led the XR development, managed the project, and designed the virtual reality environment. AX led the development of room acoustics and auralization. Chao Chia-Hsuan developed the interactive elements and ensured their seamless integration within the VR system. Giacomo Montiani developed the acoustic features within the XR environment.

References

[1] Whyte, J., "Virtual Reality and the Built Environment," *Presence*, 12(5), pp. 550–552, 2003.

[2] Portman, M. E., Natapov, A., and Fisher-Gewirtzman, D., "To go where no man has gone before: Virtual reality in architecture, landscape architecture and environmental planning," *Computers, Environment and Urban Systems*, 54, pp. 376–384, 2015.

[3] Yu, R., Gu, N., Lee, G., and Khan, A., "A Systematic Review of Architectural Design Collaboration in Immersive Virtual Environments," *Designs*, 6(5), 2022.

[4] Kuliga, S. F., Thrash, T., Dalton, R. C., and Hölscher, C., "Virtual reality as an empirical research tool — Exploring user experience in a real building and a corresponding virtual model," *Computers, Environment and Urban Systems*, 54, pp. 363–375, 2015.

[5] Volländer, M., *Auralization: Fundamentals of Acoustics, Modelling, Simulation, Algorithms and Acoustic Virtual Reality | Springer Nature Link*, Springer-Verlag, Berlin Heidelberg, 1st edition, 2008, doi:10.1007/978-3-540-48830-9.

[6] Thery, D., Boccara, V., and Katz, B. F. G., "Auralization uses in acoustical design: A survey study of acoustical consultants," *The Journal of the Acoustical Society of America*, 145(6), pp. 3446–3456, 2019.

[7] Kleiner, M., Dalenbäck, B.-I., and Svensson, P., "Auralization-An Overview," *Journal of the Audio Engineering Society*, 41(11), pp. 861–875, 1993.

[8] Lokki, T., Pätynen, J., Kuusinen, A., and Tervo, S., "Concert hall acoustics: Repertoire, listening position, and individual taste of the listeners influence the qualitative attributes and preferences," *The Journal of the Acoustical Society of America*, 140(1), pp. 551–562, 2016.

[9] Savioja, L., Huopaniemi, J., Lokki, T., and R.~Väänänen, "Creating Interactive Virtual Acoustic Environments," *Journal of the Audio Engineering Society*, 47, pp. 675–705, 1999.

[10] Tsingos, N., Gallo, E., and Drettakis, G., "Perceptual audio rendering of complex virtual environments," *ACM Trans. Graph.*, 23(3), pp. 249–258, 2004.

[11] Mo, Q., Taylor, M., Chandak, A., Lauterbach, C., Schissler, C., and Manocha, D., "Interactive GPU-based acoustic walkthrough in dynamic scenes," *Proceedings of Meetings on Acoustics*, 19, 2013.

[12] Abel, J. S., Callery, E. F., and Canfield-Dafilou, E. K., "A feedback canceling reverberator," in *Proc. Int. Conf. Digit. Audio Eff.*, pp. 100–106, 2018.

[13] Canfield-Dafilou, E. K., Callery, E. F., Abel, J. S., and Berger, J., "A method for studying interactions between music performance and rooms with real-time virtual acoustics," in *Audio Engineering Society Convention 146*, Audio Engineering Society, 2019.

[14] Raghuvanshi, N. and Gamper, H., "Interactive and Immersive Auralization," in M. Geronazzo and S. Serafin, editors, *Sonic Interactions in Virtual Environments*, pp. 77–113, Springer International Publishing, Cham, 2023.

[15] RWTH Aachen University, "VISTA - Virtual Reality and Immersive Visualization," 2025, www.vr.rwth-aachen.de/software/ViSTA.

[16] Paes, D., Arantes, E., and Irizarry, J., "Immersive environment for improving the understanding of architectural 3D models: Comparing user spatial perception between immersive and traditional virtual reality systems," *Automation in Construction*, 84, pp. 292–303, 2017.

[17] Du, J., Shi, Y., Mei, C., Quarles, J., and Yan, W., "Communication by Interaction: A Multiplayer VR Environment for Building Walkthroughs," in *Construction Research Congress 2016: Old and New Construction Technologies Converge in Historic San Juan*, pp. 2281–2290, American Society of Civil Engineers, San Juan, Puerto Rico, 2016, doi:10.1061/9780784479827.227.

[18] Hong, J. Y., He, J., Lam, B., Gupta, R., and Gan, W.-S., "Spatial Audio for Soundscape Design: Recording and Reproduction," *Applied Sciences*, 7(6), p. 627, 2017.

[19] Chung, A. and To, W. M., "Next generation soundscape design using virtual reality technologies," *Journal of the Acoustical Society of America*, 140(4), p. 3041, 2016.

[20] Pelzer, S. and Vorländer, M., "Real-Time Room Acoustics Planning," in *DAGA 2015*, pp. 1139–1142, Nürnberg, 2015.

[21] Edvaldsson, G. and Xydis, A., "Latency-free real-time auralisation in hemi-anechoic chambers," *INTER-NOISE and NOISE-CON Congress and Conference Proceedings*, 270(3), pp. 8120–8127, 2024.

[22] Scotto, F., Chia-Hsuan, C., Montiani, G., Xydis, A., Gramazio, F., and Kohler, M., "EchoXR: A Collaborative VR Framework for Spatial Acoustics in Architectural Design," in *Proceedings of the 15th International Conference on Human Interaction & Emerging Technologies*, Vienna, Austria, 2025.

[23] Dede, C., "Immersive Interfaces for Engagement and Learning," *Science*, 323(5910), pp. 66–69, 2009, doi:10.1126/science.1167311.

[24] Brown, E. and Cairns, P., "A grounded investigation of game immersion," in *CHI '04 Extended Abstracts on Human Factors in Computing Systems*, CHI EA '04, pp. 1297–1300, Association for Computing Machinery, New York, NY, USA, 2004, doi:10.1145/985921.986048.

[25] Chandrasekera, T., Yoon, S.-Y., and D'Souza, N., "Virtual environments with soundscapes: a study on immersion and effects of spatial abilities," *Environment and Planning B: Planning and Design*, 42(6), pp. 1003–1019, 2015, doi:10.1068/b130087p.

[26] Schafer, R. M., *The soundscape: our sonic environment and the tuning of the world*, Destiny Books, Rochester,Vermont, 1997.

[27] Unity Technologies, "How do I get Unity to playback a Microphone input in real time?" 2023, https://support.unity.com/hc/en-us/articles/206485253-How-do-I-get-Unity-to-playback-a-Microphone-input-in-real-time.

[28] Sanders, L. D., Zobel, B. H., Freyman, R. L., and Keen, R., "Manipulations of listeners' echo perception are reflected in event-related potentials," *The Journal of the Acoustical Society of America*, 129(1), pp. 301–309, 2011.